A weight function theory of zero order basis function interpolants and smoothers

Phillip Y. Williams

5/31 Moulden Court, Canberra, ACT 2617, Australia.

philwill@pcug.org.au, phil.williams@canberra.edu.au

April 23, 2008

I would like to thank Dr Dave Davies and the University of Canberra for their generous assistance. Also, thanks to my Masters degree supervisors Dr Markus Hegland and Dr Steve Roberts of the Centre for Mathematics and its Applications (CMA) at the Australian National University.

ABSTRACT
In this document I develop a weight function theory of zero order basis function interpolants and smoothers. In Chapter 1 the basis functions and data spaces are defined directly using weight functions. The data spaces are used to formulate the variational problems which define the interpolants and smoothers discussed in later chapters. The theory is illustrated using some standard examples of radial basis functions and a class of weight functions I will call the tensor product extended B-splines.

In Chapter 2 the theory of Chapter 1 is used to prove the pointwise convergence of the minimal norm basis function interpolant to its data function and to obtain orders of convergence. The data functions are characterized locally as Sobolev-like spaces and the results of several numerical experiments using the extended B-splines are presented.

In Chapter 3 a large class of tensor product weight functions will be introduced which I call the central difference weight functions. These weight functions are closely related to the extended B-splines and have similar properties. The theory of this document is then applied to these weight functions to obtain interpolation convergence results. To understand the theory of interpolation and smoothing it is not necessary to read this chapter.

In Chapter 4 a non-parametric variational smoothing problem will be studied using the theory of this document with special interest in its order of pointwise convergence of the smoother to its data function. This smoothing problem is the minimal norm interpolation problem stabilized by a smoothing coefficient.

In Chapter 5 a non-parametric, scalable, variational smoothing problem will be studied, again with special interest in its order of pointwise convergence to its data function. We discuss the *SmoothOperator* software (freeware) package which implements the Approximate smoother algorithm. It has a full user manual which includes several tutorials and data experiments.

Contents

0.1 Introduction

In this document I develop a weight function theory of zero order basis function interpolants and smoothers. Note that the Appendix of this document contains a list of basic notation, definitions and properties also used in this document.

This document had its genesis in the development of a *scalable algorithm* for *Data Mining* applications. Data Mining is the extraction of complex information from large databases, often having tens of millions of records. Scalability means that the time of execution is linearly dependent on the number of records processed and this is necessary for the algorithms to have practical execution times. One approach is to develop additive regression models and these require the approximation of large numbers of data points by surfaces. Here one is concerned with approximating data by surfaces of the form $y = f(x)$, where $x \in \mathbb{R}^d$, $y \in \mathbb{R}$ and d is any dimension. Smoothing algorithms are one way of approximating surfaces and in particular we have decided to use a class of non-parametric smoothers called basis function smoothers, which solve a variational smoothing problem over a semi-Hilbert space of continuous *data functions* and express the solution in terms of a single *basis function*.

I started my Masters degree (supervised by Dr. Markus Hegland and Dr. Steve Roberts at the ANU, Canberra, Australia) searching for a scalable basis function smoothing algorithm and had the good fortune to devise such an algorithm (unpublished) by approximating, on a regular grid, the convolution in Definition 38 of the space J_G in Dyn's review article [3]. In this document I develop some theoretical tools to construct and analyze this algorithm for the case of positive order basis functions. In the document Williams [13] the zero order case is studied. This theoretical approach applies in any dimension but the smoothing algorithm is only practical up to about three dimensions. In higher dimensions the matrices are too large to put into computer memory.

Dr. Hegland was particularly interested in using the tensor product hat (triangular) function as a basis function. At about this time we had a visit by the late Professor Will Light who showed me his paper [8] which defined basis functions in terms of weight functions using the Fourier transform. Light and Wayne's weight function properties were designed for positive order basis functions which excluded the tensor product hat function. They were designed for the well-known 'classical' radial weight functions. I therefore developed a version of his theory designed to generate zero order basis functions, including both tensor product and radial types. This theory is developed in Chapter 1 and requires that the basis functions have Fourier transforms which can take zero values outside the origin since this is a property of hat functions.

Chapter by chapter in brief:

- **Chapter** 1: The goal of this chapter is to extend the theoretical work of Light and Wayne in [8] to allow classes of weight functions analogous to the positive order weight functions developed in Chapter 1 of Williams [13]. Both the inner product *data space* and the *basis functions* are defined in terms of the weight function. We then prove the completeness and smoothness properties of the data space as well as the continuity and positive definiteness properties of the basis function. We use as examples the radial weight functions: the thin-plate splines, the Gaussian and the Sobolev spline. Special attention is paid to a class of tensor product weight functions we call the *extended B-splines* which are the convolutions of hat functions.

- **Chapter** 2 studies the *minimum seminorm interpolation problem*. Topics include the existence and uniqueness of the basis function solution, a matrix equation for the solution and the pointwise convergence to the interpolant to its data function. Pointwise order of convergence results are derived using two techniques. One approach uses Lagrange interpolation techniques and the second does not. Numerical experiments were undertaken using the extended B-splines and special data functions were constructed for these experiments. This led to the extended B-splines data functions being characterized locally as Sobolev-like spaces.

- **Chapter** 3: This chapter introduces a large new class of weight functions which I call the *central difference weight functions* because they are based on central difference operators. They are closely related to the *extended B-splines*. The central difference basis functions are calculated and smoothness estimates and upper and lower bounds are derived for the weight functions and the basis functions. Like the extended B-splines their data functions are characterized locally as Sobolev-like spaces. Pointwise convergence results are derived for the basis function interpolant using results from Chapter 2. These give orders of convergence identical to those obtained for the extended B-splines in. We do not present the results of any numerical experiments.

- **Chapter** 4 studies the well-known *Exact smoother problem* (our terminology) which stabilizes the interpolant by adding a smoothing coefficient to the seminorm functional. Topics include the existence and uniqueness of the basis function solution, a matrix equation for the solution and the pointwise convergence of the solution to its data function. Orders of convergence (also called error orders) are derived which, in special sense, are the same as those obtained for the interpolant.

- **Chapter** 5 studies the scalable *Approximate smoother* (our terminology) which turns out to have a discretization process similar to that described in Garcke and Griebel [4]. Topics include the existence and uniqueness of the smoother, matrix equations for the smoother, and the pointwise convergence of this smoother to the Exact smoother and to its data function. Orders of pointwise convergence are derived for the convergence of the Approximate to the Exact smoother which are then combined with the Exact smoother error formulas of Chapter 4 to obtain error orders for the Approximate smoother.

Chapter by chapter in more detail:

0.1.1 Chapter 1 Weight functions, data spaces and basis functions

In his paper [2] Duchon studied positive order basis function interpolation. Basically he began by generalizing the functional $\int\limits_{\mathbb{R}^d} \sum\limits_{|\alpha|=m} |D^\alpha u|^2$ using a positive weight function w and the Fourier transform to obtain $\int\limits_{\mathbb{R}^d} w \sum\limits_{|\alpha|=m} \left|\widehat{D^m u}\right|^2$, and then devoted the paper to studying the special case $w = |\cdot|^{2s}$. In [8] Light and Wayne carried on the study of the positive order interpolation problem where Duchon left off. Their weight function theory used a strictly positive weight function to directly define the basis functions and Hilbert spaces.

In this document I adapt the work of Light and Wayne to the study the much simpler zero order problem. Here the weight function is only assumed to be a.e. positive and the conditions placed on the weight functions are expressed in terms of integrals. The weaker condition on the weight function allows for basis functions whose Fourier transforms have zeros e.g. the hat or triangular function. Happily, the zero order theory presented here only requires simple L^1 function space theory, the positive order results being considerably more complex - see Williams [13].

The theory is illustrated using the standard radial basis functions: the *shifted thin-plate splines*, the *Gaussian* and the *Sobolev splines* but I am especially interested in basis functions that are the tensor product of the derivatives of hat (triangle) functions, denoted Λ, which I call the *extended B-splines*. However, as mentioned previously, the Fourier transform of these basis functions has zero values on a set of measure zero and the theory has to allow for this. Thus I begin by using a class of a.e. positive and continuous weight functions w satisfying (property W2): $\int_{\mathbb{R}^d} \frac{|x|^{2\lambda}}{w(x)} dx < \infty$ for $0 \leq \lambda \leq \kappa$, to define the zero order data space

$$X_w^0 = \left\{ f \in S' : \widehat{f} \in L^1_{loc} \text{ } and \text{ } \sqrt{w}\widehat{f} \in L^2 \right\}.$$

This turns out to be a *reproducing kernel Hilbert space* of continuous functions with norm $\|f\|_{w,0}$ and inner product $(f,g)_{w,0} = \int w \widehat{f}\overline{\widehat{g}}$. It is shown, for example, that $X_w^0 \hookrightarrow C^{(\lfloor \kappa \rfloor)}$ and that $S \cap X_w^0$ is dense in X_w^0.

The weight function is then used to define a zero order basis function G by the simple Fourier transform formula $\widehat{G} = \frac{1}{w}$. The smoothness obtained is $G \in C_B^{(\lfloor 2\kappa \rfloor)}$ and several results are proved regarding the tensor products of weight functions and basis functions, as well as the products and convolutions of basis functions.

The *extended B-splines* are discussed in detail. It turns out that the corresponding basis functions are the derivatives of convolutions of hat (triangle) functions and hence their name.

0.1.2 Chapter 2 The minimal norm interpolant

In this chapter the minimal norm interpolation problem is solved and several pointwise convergence results are derived. These are illustrated with numerical results obtained using several extended B-spline basis functions and special classes of data functions. In more detail, the functions from the Hilbert data space X_w^0 are used to define the standard minimal norm interpolation problem with independent data $X = \left\{x^{(i)}\right\}_{i=1}^N$ and dependent data $y = \{y_i\}_{i=1}^N$, the latter being obtained by evaluating a data function at X. This problem is then shown to have a unique basis function solution of the form $\sum\limits_{i=1}^N \alpha_i G\left(\cdot - x^{(i)}\right)$ with $\alpha_i \in \mathbb{C}$. The standard matrix equation for the α_i is then derived.

We will then consider several categories of estimates for the pointwise convergence of the interpolant to its data function when the data is confined to a bounded data region. In all these results the order of convergence appears as the power of the radius of the largest ball in the data region that can be fitted between the X data points. The convergence results of this document can be divided into those which use Lagrange interpolation theory, and thus assume unisolvent data X (Definition 66), and those which do not. The non-unisolvency proofs are *much* simpler than the unisolvency-based results and if the data functions are chosen appropriately the constants can be calculated. However, the maximum order of

convergence obtained is 1, no matter how large the parameter κ. On the other hand, the constants for the unisolvency-based results are difficult to calculate but the order of convergence is equal to the value of $\max\lfloor\kappa\rfloor$.

Non-unisolvent data: Type 1 interpolation error estimates Suppose that the weight function has property W2 for some $\kappa \geq 0$, that the data X is contained in a closed, bounded, infinite data set K, and that the basis function G satisfies the estimate

$$|G(0) - \operatorname{Re} G(x)| \leq C_G |x|^{2s}, \quad |x| < h_G, \tag{1}$$

for some $h_G > 0$. Then in Theorem 59 it is shown that the interpolant $\mathcal{I}_X f$ of a data function $f \in X_w^0$ satisfies

$$|f(x) - \mathcal{I}_X f(x)| \leq k_G \|f\|_{w,0} (h_{X,K})^s, \quad x \in K, \tag{2}$$

when $h_{X,K} = \max\limits_{x \in K} \operatorname{dist}(x, X) \leq h_G$ and $k_G = (2\pi)^{-\frac{d}{4}} \sqrt{2C_G}$. This implies an order of convergence of at least s.

These results are summarized in Table 2.1 (a copy of Table 2.1).

Interpolant error estimates Non-unisolvent **Type 1** estimates: $\kappa \geq 0$, $k_G = (2\pi)^{-d/4}\sqrt{2C_G}$.				
Weight function	Parameter constraints	Converg. order s	C_G	h_G
Sobolev splines	$v - \frac{d}{2} = 1$	$\frac{1}{2}$	$\frac{\|\rho K_0(\rho)\|_\infty}{2^{v-1}\Gamma(v)}$	∞
$(v > d/2)$	$v - \frac{d}{2} \neq 1$	1	$\frac{\|D^2 K_{v-d/2}\|_\infty}{2^v \Gamma(v)}$	"
Shifted thin-plate $(-d/2 < v < 0)$	-	1	eq. (2.26)	∞
Gaussian	-	1	$2e^{-3/2}$	∞
Extended B-spline	-	$\frac{1}{2}$	$G_1(0)^{d-1} \|DG_1\|_\infty \sqrt{d}$	∞

TABLE 1.

Non-unisolvent data: Type 2 interpolation error estimates Here we avoid making any assumptions about G and instead assume $\kappa \geq 1$ and use the properties of the Riesz representer R_x of the evaluation functional $f \to f(x)$. A consequence of this approach is the estimate (Theorem 65): when $h_{X,K} < \infty$

$$|f(x) - \mathcal{I}_X f(x)| \leq k_G \|f\|_{w,0} h_{X,K}, \quad x \in K, \ f \in X_w^0, \tag{3}$$

where

$$k_G = (2\pi)^{-d/4} \sqrt{-(\nabla^2 G)(0)}.$$

so the order of convergence is **always** at least 1. Table 2 (a copy of Table 2.2) summarizes the results for the weight function examples:

Unisolvent data - interpolation error estimates Here X is assumed to have a unisolvent subset of order $\max\lfloor\kappa\rfloor$ and is contained in an open, bounded data region Ω. In this case it follows from Theorem 74 that there exist constants $h_{\Omega,\kappa}, k_G > 0$ such that

$$|f(x) - \mathcal{I}_X f(x)| \leq k_G \|f\|_{w,0} (h_{X,\Omega})^{\lfloor\kappa\rfloor}, \quad x \in \overline{\Omega}, \ f \in X_w^0, \tag{4}$$

when $h_{X,\Omega} = \sup\limits_{\omega \in \Omega} \operatorname{dist}(\omega, X) < h_{\Omega,\kappa}$. Table 3 (a copy of Table 2.3) summarizes our results:

Numerical results are only presented for the non-unisolvent data cases. Numerical results are presented which illustrate the convergence of the interpolant to its data function. We will only be interested in the convergence of the interpolant to it's data function and not in the algorithm's performance as an interpolant. Only the extended B-splines will be considered and we will also restrict ourselves to one dimension so that the data density parameter $h_{X,\Omega}$ can be easily calculated.

If we can calculate the data function norm we can calculate the error estimate. All the extended B-spline basis weight functions have a power of $\sin^2 x$ in the denominator and so we have derived special classes

Interpolant error estimates Non-unisolvent **Type 2** estimate: $\kappa \geq 1$			
Weight function	Parameter constraints	Converg. order	$(2\pi)^{d/4}\, k_G/\sqrt{d}$
Sobolev splines	$v - \frac{d}{2} \geq 2$	1	$\sqrt{\frac{\Gamma(v-d/2-1)}{2^{d/2+1}\Gamma(v)}}$
$(v > d/2)$	$1 < v - \frac{d}{2} < 2$	"	$\sqrt{\frac{\Gamma(v-d/2-1)}{2^{2v-d/2-3}\Gamma(v)}}$
Shifted thin-plate $(-d/2 < v < 0)$	-	1	$\sqrt{-2v}$
Gaussian	-	1	$\sqrt{2}$
Extended B-spline $(1 \leq n \leq l)$	$n \geq 2$	1	$\sqrt{-G_1\,(0)^{d-1}\, D^2 G_1\,(0)}$

TABLE 2.

Interpolant error estimates Unisolvent data: $\kappa \geq 1$		
Weight function	Parameter constraints	Convergence orders ($\lfloor \kappa \rfloor$)
Sobolev splines	$v - \frac{d}{2} = 2, 3, 4, \ldots$	$\lfloor v - d/2 \rfloor - 1$
$(v > d/2)$	$v - \frac{d}{2} > 1,\ v - \frac{d}{2} \notin \mathbb{Z}_+$	$\lfloor v - d/2 \rfloor$
Shifted thin-plate spline	-	$2, 3, 4, \ldots$
Gaussian	-	$2, 3, 4, \ldots$
Extended B-spline $(1 \leq n \leq l)$	$n \geq 2$	$n - 1$

TABLE 3.

of data functions for which the data function norm can be calculated. The derivation of these special classes of data functions led to the characterization of the restriction spaces $X_w^0\,(\Omega)$ for various class of weight function. For example, Theorem 90 shows that the restrictions of the B-spline data functions are a member of the class of *Sobolev-like spaces*:

$$H_s^m\,(\Omega) = \left\{ u \in L^2\,(\Omega) : D^\alpha u \in L^2\,(\Omega)\ \ for\ \alpha_i \leq m,\ 0 \leq i \leq m \right\}, \quad m = 1, 2, 3, \ldots \tag{5}$$

It is easy to construct functions that are in these spaces.

As expected interpolant instability is evident and because our error estimates assume an infinite precision we filter the error to remove spikes which are a manifestation of the instability.

0.1.3 Chapter 3 The central difference weight functions

To understand the theory of interpolants and smoothers presented in this document it is not necessary to read this chapter which introduces the *central difference weight functions*. However, we note that Section 3.5 contains local data space results specifically designed for tensor product weight functions and that central difference weight functions are used as examples in the chapters dealing with smoothers.

This chapter introduces a large new class of weight functions based on central difference operators. The basis functions are calculated and smoothness estimates and upper and lower bounds are derived for the weight functions and the basis functions.

Like the extended B-splines, their data functions are characterized locally as Sobolev-like spaces.

Pointwise convergence results are derived for the basis function interpolant using results from Chapter 2. These give orders of convergence identical to those obtained for the extended B-splines in. This chapter presents no numerical interpolation experiments.

This is a large class of weight functions and I have only calculated a couple of basis functions. The calculations can be tedious so one really needs to be able to have theoretical results which enable you to tell beforehand that the basis function has some especially good properties regarding interpolation - I have no results to offer in this regard.

The central difference weight functions are defined as follows: Suppose $q \in L^1\left(\mathbb{R}^1\right)$, $q \neq 0$, $q(\xi) \geq 0$ and $l \geq n \geq 1$ are integers. The univariate *central difference weight function* with parameters n, l is defined by

$$w(\xi) = \frac{\xi^{2n}}{\Delta_{2l}\widehat{q}(\xi)}, \quad \xi \in \mathbb{R}^1,$$

where Δ_{2l} is central difference operator

$$\Delta_{2l} f(\xi) = \sum_{k=-l}^{l} (-1)^k \tbinom{2l}{k+l} f(-k\xi), \quad l = 1, 2, 3, \ldots, \ \xi \in \mathbb{R}^1.$$

For example, $\Delta_2 f(\xi) = -(f(\xi) - 2f(0) + f(-\xi))$. The multivariate central difference weight function is defined by tensor product.

It is shown in Theorem 98 that w belongs to the class of zero order weight functions introduced in Chapter 1 for some κ iff $\int\limits_{|\xi| \geq R} |\xi|^{2n-1} q(\xi)\, d\xi < \infty$ for some $R \geq 0$. Here κ satisfies $\kappa + 1/2 < n$. The central difference weight functions are closely related to the *extended B-splines* defined by 1.20 and a discussion of their genesis is given in Subsection 3.2.1.

Various bounds are derived for w. For example, in Corollary 109 it is shown that if $\int_{|t| \geq R} t^{2l} q(t)\, dt < \infty$ then for any $r > 0$ there exist constants $c_r, c_r', k_r, k_r' > 0$ such that

$$\begin{gathered} k_r \xi^{2n} \leq w(\xi) \leq k_r' \xi^{2n}, \quad |\xi| \geq r, \\ \frac{c_r}{\xi^{2(l-n)}} \leq w(\xi) \leq \frac{c_r'}{\xi^{2(l-n)}}, \quad |\xi| \leq r. \end{gathered}$$

By Theorems 111 and 115 the univariate *central difference basis function* is

$$G(s) = (-1)^{(l-n)} \int\limits_{\mathbb{R}^1} \left(D^{2(l-n)} \left(*\Lambda\right)^l\right) \left(\frac{s}{t}\right) |t|^{2n-1} q(t)\, dt, \quad s \in \mathbb{R}^1,$$

with $G \in C_B^{(2n-2)}$ and $D^{2n-1}G$ bounded in $C^{(2n-1)}(\mathbb{R}\backslash 0)$. The multivariate basis function is defined as a tensor product. For $k \leq 2n-2$, $D^k G$ is uniformly Lipschitz continuous of order 1 (Theorem 118).

In Section 3.5 we will characterize the *data function space* locally as a Sobolev-like space 5. A localization result specifically for tensor product weight functions is derived in Theorem 123. This supplies important information about the data functions and makes it easy to choose data functions for numerical experiments.

Results for the pointwise convergence of the interpolant to its data function on a bounded data domain are derived using results from Chapter 1. These results are summarized here in Table 4 (a copy of Table 3.1) and give pointwise orders of convergence identical to those obtained for the extended B-splines.

Interpolant convergence to data function - central difference weight function.					
Estimate name	Unisolv. data?	Parameter constraints	Converg. order	$(2\pi)^{d/4} k_G$	h_G
Type 1	no	q b'nded if $n=1$ else $n \geq 2$	1/2	$\sqrt{2G_1(0)^{d-1} \lVert DG_1 \rVert_\infty} \sqrt[4]{d}$	∞
Type 2	no	$n \geq 2$	1	$\sqrt{-G_1(0)^{d-1} D^2 G_1(0)} \sqrt{d}$	∞
Unisolvent	yes	$n \geq 2$	$n-1$	-	∞

TABLE 4.

This chapter does not present the results of any numerical experiments.

0.1.4 Chapter 4 The Exact smoother

We call this well-known basis function smoother the *Exact smoother* because the smoother studied in the next chapter approximates it. We will assume the basis function is real-valued. The Exact smoother

minimizes the functional

$$\rho \|f\|_{w,0}^2 + \frac{1}{N}\sum_{i=1}^{N} \left| f(x^{(i)}) - y_i \right|^2, \quad f \in X_w^0, \tag{6}$$

over the data (function) space X_w^0 where $\rho > 0$ is termed the smoothing coefficient. It is shown that this problem, like the interpolation problem, has a unique basis function solution in the space $W_{G,X}$. The finite dimensionality of the solution allows us to derive matrix equations for the coefficients α_i of the data-translated basis functions.

Three estimates are derived for the order of the pointwise error of the Exact smoother w.r.t. its data function: two estimates do not assume unisolvent X data and one does. When $\rho = 0$ these estimates correspond to interpolation results of Subsection 2.5.2 and the Exact smoother convergence orders and the constants are the same as those for the interpolation case which are given in the interpolation tables 1, 2, 3 and 4.

$\boxed{\text{Non-unisolvent data: Type 1 error estimates}}$ When the weight function has property W2 for $\kappa \geq 0$ it is assumed that the basis function satisfies an inequality of the form 1 and that the data region is a closed, bounded, infinite set K. In this case it is shown in Corollary 153 that the Exact smoother s_e of the data function f satisfies the error estimate

$$|f(x) - s_e(x)| \leq \|f\|_{w,0} \left(\sqrt{\rho N} + k_G (h_{X,K})^s \right), \quad x \in K, \tag{7}$$

when $h_{X,K} = \max\limits_{x \in K} \operatorname{dist}(x, X) \leq h_G$ and $k_G = (2\pi)^{-d/4} \sqrt{2C_G}$.

$\boxed{\text{Non-unisolvent data: Type 2 error estimates}}$ If it only assumed that $\kappa \geq 1$ then it is shown in Theorem 156 that

$$|f(x) - s_e(x)| \leq \|f\|_{w,0} \left(\sqrt{\rho N} + k_G h_{X,K} \right), \quad x \in K, \tag{8}$$

when $h_{X,K} < \infty$ and $k_G = (2\pi)^{-d/4} \sqrt{-(\nabla^2 G)(0)}\sqrt{d}$.

$\boxed{\text{Unisolvent data error estimates}}$ On the other hand, suppose the weight function has property W2 for some parameter $\kappa \geq 1$, the independent data X contains a unisolvent set of order $\lfloor \kappa \rfloor$ and X is contained in an open, bounded data region Ω. Then using results from the Lagrange theory of interpolation we show in Theorem 160 that there exist constants $K'_{\Omega,m}, k_G > 0$ such that

$$|f(x) - s_e(x)| \leq \|f\|_{w,0} \left(K'_{\Omega,m} \sqrt{\rho N} + k_G (h_{X,\Omega})^{\lfloor \kappa \rfloor} \right), \quad x \in \overline{\Omega}, \tag{9}$$

when $h_{X,\Omega} = \sup\limits_{x \in \Omega} \operatorname{dist}(x, X) \leq h_G$.

These theoretical results will be illustrated using the weight function examples from the interpolation chapter, namely the radial *shifted thin-plate splines*, *Gaussian* and *Sobolev splines* and the tensor product *extended B-splines*. We will also use the *central difference* weight functions from Chapter 3.

Numerical results are only presented for the non-unisolvent data cases. Numeric experiments are carried out using the same 1-dimensional B-splines and data functions that were used for the interpolants. We restrict ourselves to one dimension so that the data density parameters $h_{X,\Omega}$ and $h_{X,K}$ can be easily calculated.

0.1.5 Chapter 5 The Approximate smoother

We call this smoother the *Approximate smoother* because it approximates the Exact smoother. This is a non-parametric, scalable smoother. Here *scalable* means the numeric effort to calculate the Approximate smoother depends linearly on the number of data points. We assume the basis function is real-valued.

Two different approaches will be taken to defining the Approximate smoother, and both involve formulating the smoother as the solution of a variational problem. One of these problems will involve minimizing the Exact smoother functional 6 over $W_{G,X'}$ where $X' = \{x'_i\}_{i=1}^{N'}$ is an arbitrary set of distinct points in $\mathbb{R}^d$. The other, equivalent problem, involves finding the function in $W_{G,X'}$ which is nearest

to Exact smoother s_e w.r.t. the norm $\|\cdot\|_{w,0}$. If

$$s_a(x) = \sum_{i=1}^{N'} \alpha'_i R_{x'_i}(x) = (2\pi)^{-d/2} \sum_{i=1}^{N'} \alpha'_i G(x - x'_i),$$

denotes the Approximate smoother and y is the dependent data then solving the second problem yields

$$s_a = \mathcal{I}_{X'} s_e,$$

which implies the matrix equation

$$\left(N\rho R_{X',X'} + R_{X,X'}^T R_{X,X'}\right)\alpha' = R_{X,X'}^T y,$$

where $R_{X,X'} = \left(R_{x'_j}\left(x^{(i)}\right)\right)$. The size of the Approximate smoother matrix is $N' \times N'$ which is independent of the number of data points and suggests scalability.

The error estimates for the pointwise convergence of the Approximate smoother to its data function $f \in X_w^0$ are based on the simple triangle inequality

$$|f(x) - s_a(x)| \le |f(x) - s_e(x)| + |s_e(x) - s_a(x)|,$$

and so Section 5.6 will be devoted to estimating $|s_e(x) - s_a(x)|$.

As with the minimal interpolant and the Exact smoother, we will obtain estimates that assume unisolvent data sets as well as the Type 1 and Type 2 estimates that do not involve unisolvency. The Approximate smoother convergence orders and the constants are the same as those for the interpolation case which are given in the interpolation tables 1, 2, 3 and 4.

Non-unisolvent data: Type 1 error estimates No *a priori* assumption is made concerning the weight function parameter κ but it will be assumed that the basis function satisfies an inequality of the form 1. For example, if it is assumed that the data region K is closed, bounded and infinite then Theorem 189 establishes that

$$|s_e(x) - s_a(x)| \le \|f\|_{w,0}\, k_G \left(h_{X',K}\right)^s, \quad x \in K, \tag{10}$$

and Theorem 206 shows that

$$|f(x) - s_a(x)| \le \|f\|_{w,0} \left(\sqrt{\rho N} + k_G \left(h_{X,K}\right)^s + k_G \left(h_{X',K}\right)^s\right), \quad x \in K, \tag{11}$$

when $h_{X,K} = \max\limits_{x \in K} \operatorname{dist}(x, X) \le h_G$ and $h_{X',K} = \max\limits_{x \in K} \operatorname{dist}(x, X') \le h_G$.

Non-unisolvent data: Type 2 error estimates If it only assumed that $\kappa \ge 1$ then by Theorem 196

$$|s_e(x) - s_a(x)| \le \|f\|_{w,0}\, k_G \left(h_{X',K}\right)^s, \quad x \in \mathbb{R}^d, \tag{12}$$

and by Theorem 208

$$|f(x) - s_a(x)| \le \|f\|_{w,0} \left(\sqrt{\rho N} + k_G h_{X,K} + k_G h_{X',K}\right), \quad x \in \mathbb{R}^d, \tag{13}$$

where $k_G = (2\pi)^{-\frac{d}{4}} \sqrt{-\left(\nabla^2 G\right)(0)}\sqrt{d}$.

Unisolvent data error estimates If X is a unisolvent set of order $m \ge 1$ contained in an open, bounded data region Ω then by Theorem 200

$$|s_e(x) - s_a(x)| \le \|f\|_{w,0}\, k_G \left(h_{X',\Omega}\right)^m, \quad x \in \overline{\Omega}, \tag{14}$$

and by Theorem 211

$$|f(x) - s_a(x)| \le \|f\|_{w,0} \left(K'_{\Omega,m}\sqrt{\rho N} + k_G \left(h_{X,K}\right)^m + k_G \left(h_{X',K}\right)^m\right), \quad x \in \overline{\Omega}, \tag{15}$$

for some constants $K'_{\Omega,m}, k_G > 0$. We say the orders of convergence are at least m.

These theoretical results will be illustrated using the weight function examples from the interpolation chapter, namely the radial *shifted thin-plate splines*, *Gaussian* and *Sobolev splines* and the tensor product *extended B-splines*. We will also use the *central difference* weight functions from Chapter 3.

Numerical results are only presented for the non-unisolvent data cases. Numeric experiments are carried out using the same 1-dimensional B-splines and data functions that were used for the interpolants. We restrict ourselves to one dimension so that the data density parameters $h_{X,\Omega}$ and $h_{X,K}$ can be easily calculated.

We discuss the *SmoothOperator* software (freeware) package which implements the Approximate smoother algorithm. It has a full user manual which describe several tutorials and data experiments.

1

Weight functions, data spaces and basis functions

1.1 Introduction

In [8] Light and Wayne developed a weight function theory of positive order basis function interpolation in which the weight function directly generated both the basis function and the Hilbert space of continuous data functions. We intend to do the same thing here for the zero order case.

We start by introducing the theory of weight functions using as examples the radial basis functions: the shifted thin-plate splines, the Gaussian, the Sobolev splines as well as the tensor product extended B-spline weight functions. Results are also proved regarding the tensor products and products of weight functions.

The weight functions are then used to define both the Hilbert data spaces and the basis functions which will be used to formulate and solve the variational interpolation and smoothing problems discussed in later Chapters. Various continuity results are derived for the basis functions and data space functions. The data space is shown to be a reproducing kernel Hilbert space but instead of using the reproducing kernel we will use the Riesz representers of the evaluation functionals $f \to D^{\alpha} f(x)$ because the kernel is only defined for $\alpha = 0$.

1.2 The weight functions

1.2.1 Motivation for the weight function properties

We want to modify Light and Wayne's weight function properties and basis function definition so that the tensor product hat functions are basis functions of zero order generated by weight functions. We also hope to use weight and basis functions that require only simple L^1 Fourier transform results and avoid subspaces of the tempered distributions.

Light and Wayne defined their weight function w to have the three properties:

A3.1 w is continuous and positive outside the origin.

A3.2 $\frac{1}{w} \in L^1_{loc}$.

A3.3 For some $R > 0$ and μ real, $\frac{1}{w(x)} \leq c\,|x|^{\mu}$ for $|x| > R$.

These properties imply $\frac{1}{w} \in S'$ and Light and Wayne's definition of a tempered *basis distribution* $G \in S'$ of order $k \geq 1$ generated by w (see Section 3 of [8]) imply that as distributions

$$\widehat{G} = \frac{1}{|\xi|^{2k}\, w}, \quad on\ \mathbb{R}^d \backslash 0.$$

Now, as mentioned above, we want to define a basis function so the multivariate hat function is a basis function of order zero. The one dimensional hat function will be denoted by Λ and is given by

$$\Lambda(x) = \begin{cases} 0, & |x| > 1, \quad x \in \mathbb{R}, \\ 1 - |x|, & |x| \le 1, \quad x \in \mathbb{R}. \end{cases} \tag{1.1}$$

In higher dimensions the hat function is defined as the tensor product

$$\Lambda(x) = \prod_{i=1}^{d} \Lambda(x_i), \quad x \in \mathbb{R}^d. \tag{1.2}$$

It is well known that

$$\widehat{\Lambda}(\xi) = (2\pi)^{-1/2} \left(\frac{\sin(\xi/2)}{\xi/2} \right)^2, \quad \xi \in \mathbb{R}^1, \tag{1.3}$$

so that in higher dimensions

$$\widehat{\Lambda}(\xi) = \prod_{i=1}^{d} \widehat{\Lambda}(\xi_i) = (2\pi)^{-d/2} \prod_{i=1}^{d} \left(\frac{\sin \xi_i/2}{\xi_i/2} \right)^2, \quad \xi \in \mathbb{R}^d. \tag{1.4}$$

In one dimension, if Λ is to be basis function of order zero then we must have $\widehat{\Lambda}(\xi) = \frac{1}{w(\xi)}$ i.e.

$$w(\xi) = (2\pi)^{d/2} \prod_{i=1}^{d} \left(\frac{\xi_i/2}{\sin(\xi_i/2)} \right)^2, \quad \xi \in \mathbb{R}^d, \tag{1.5}$$

and w has discontinuities outside the origin, which violates weight function property A3.1. Denoting the set of discontinuities by $\mathcal{A}$ we have that $\mathcal{A}$ is the union of hyperplanes

$$\mathcal{A} = \bigcup_{i=1}^{d} \{\xi : \xi_i \in 2\pi\mathbb{Z}\backslash 0\}. \tag{1.6}$$

and this is a closed set of measure zero. This will become weight function property W1 in Definition 1 below.

Additionally, from equation 1.4 we have that $\widehat{\Lambda} = \frac{1}{w} \in L^1$ and we want to only use L^1 Fourier transform results. So we will say that G is a basis function of order 0 if $\frac{1}{w} \in L^1$ and $\widehat{G} = \frac{1}{w}$. A well-known L^1 result then implies $G \in C_B^{(0)} \subset S'$ and so G is a function. That $\frac{1}{w} \in L^1$ is ensured by property W2 in Definition 1. Note that our definition of a basis function gives a unique basis function whereas in Light and Wayne a basis function of higher order k is unique modulo a polynomial of order $\le k$.

Finally we will incorporate the parameter κ into property W2 to increase the smoothness of the basis function and to allow a larger choice of weight functions. The parameter κ will be allowed to take non-integer values.

1.2.2 The weight function properties

Definition 1 ***Weight functions with real parameter*** $\kappa \ge 0$

A weight function's properties are defined with reference to a set $\mathcal{A} \subset \mathbb{R}^d$ which is closed and has measure zero. The introduction of $\mathcal{A}$ is motivated directly by the properties of the Fourier transform of the hat function.

In this document a weight function w is a mapping $w : \mathbb{R}^d \to \mathbb{R}$ which has at least the property W1:

W1 *There exists a closed set $\mathcal{A}$ with measure zero such that w is continuous and positive outside $\mathcal{A}$ i.e. $w \in C^{(0)}(\mathbb{R}^d \backslash \mathcal{A})$ and $w > 0$ on $\mathbb{R}^d \backslash \mathcal{A}$.*

It may also have property W2:

W2 *For some real number $\kappa \ge 0$*

$$\int \frac{|x|^{2\lambda}}{w(x)} dx < \infty, \quad 0 \le \lambda \le \kappa. \tag{1.7}$$

Remark 2 *1. In Section 1.3 property W1 is used to define the inner product space of distributions X_w^0. Property W1 is needed to ensure that X_w^0 is not empty - see Theorem 9. It is not enough just to have a weight function that is positive a.e.*

2. *Property W1 allows basis functions to have Fourier transforms with zeros outside the origin e.g. the tensor product hat function.*

3. *I have chosen the integral form for W2 because it is more general than the form A3.3 of Light and Wayne.*

4. *Property W2 allows the definition of a continuous basis function of order zero.*

5. *The parameter κ of property W2 could be called the smoothness parameter because in Theorem 14 it is shown that $X_w^0 \subset C_B^{(\lfloor\kappa\rfloor)}$ and in Theorem 24 it is shown that the basis functions are in $C_B^{(\lfloor 2\kappa\rfloor)}$. The allowance of non-integer values of κ sometimes permits an extra degree of differentiability to be estimated e.g. The smoothness of Sobolev spline basis functions in Subsubsection 1.4.2.*

6. *Scaled (or dilated) weight functions are also weight functions with the same parameter κ.*

7. *Properties such as: $\int \frac{|x|^{2\lambda}}{w(x)^2} dx < \infty$ for $0 \leq \lambda \leq \kappa$, and $\frac{|x|^\mu}{w(x)}$ bounded for $\mu \leq v$, might be investigated in order to obtain Sobolev space results.*

The following result gives two equivalent criteria for the weight function property W2.

Theorem 3 *The the following criteria are equivalent to weight function property W2.*

1. $\int \frac{x^{2\beta}|x|^{\kappa-\lfloor\kappa\rfloor}}{w(x)} dx < \infty, \quad |\beta| \leq \lfloor\kappa\rfloor.$

2. $\frac{1}{w} \in L^1_{loc}$ *and,* $\int\limits_{|\cdot|\geq R} \frac{|\cdot|^{2\kappa}}{w} < \infty$ *for any* $R \geq 0$.

Proof. Part 1 is easily proved using the identity $|x|^{2k} = \sum\limits_{|\beta|=k} \frac{k!}{\beta!} x^{2\beta}$, and clearly W2 implies part 2. Finally, if part 2 holds for $0 \leq \lambda \leq \kappa$ and any $R \geq 0$

$$\begin{aligned}\int \frac{|\cdot|^{2\lambda}}{w} = \int\limits_{|\cdot|\leq R} \frac{|\cdot|^{2\lambda}}{w} + \int\limits_{|\cdot|\geq R} \frac{|\cdot|^{2\lambda}}{w} &\leq R^{2\lambda} \int\limits_{|\cdot|\leq R} \frac{1}{w} + \int\limits_{|\cdot|\geq R} \frac{1}{|\cdot|^{2(\kappa-\lambda)}} \frac{|\cdot|^{2\kappa}}{w} \\ &\leq R^{2\lambda} \int\limits_{|\cdot|\leq R} \frac{1}{w} + \int\limits_{|\cdot|\geq R} \frac{1}{R^{2(\kappa-\lambda)}} \frac{|\cdot|^{2\kappa}}{w} \\ &< \infty,\end{aligned}$$

and so property W2 is satisfied. ■

1.2.3 Examples: radial weight functions

Shifted thin-plate splines

From equations 25, 26 and 27 of Dyn [3] the shifted thin-plate spline functions

$$H(x) = \begin{cases} \frac{(-1)^{v+1}}{2} \left(1+|x|^2\right)^v \log\left(1+|x|^2\right), & v = 1,2,3,\ldots, \\ (-1)^{\lceil v\rceil} \left(1+|x|^2\right)^v, & v > -d/2,\ v \neq 0,1,2,3\ldots, \end{cases} \tag{1.8}$$

have the distribution Fourier transforms

$$\widehat{H}(\xi) = \widetilde{e}(v)\, \widetilde{K}_{v+d/2}(|\xi|)\, |\xi|^{-2v-d} \text{ on } \mathbb{R}^d\backslash 0, \quad v > -d/2,$$

where $\widetilde{e}(v) > 0$, $\widetilde{K}_\lambda(t) = t^\lambda K_\lambda(t)$, $t \geq 0, \lambda > 0$, and K_λ is called a modified Bessel function or MacDonald's function. Here $\widetilde{K}_\lambda$ has the properties

$$\widetilde{K}_\lambda \in C^{(0)}\left(\mathbb{R}^1\right), \ \lambda > \frac{k}{2}; \ \widetilde{K}_\lambda(t) > 0; \ \lim_{t\to\infty} \widetilde{K}_\lambda(t) = 0 \textit{ exponentially.} \tag{1.9}$$

Also, for a given $\lambda > 0$, it is well-known that there exist positive constants $1 < b_\lambda$, $0 < c_\lambda \leq c'_\lambda$ such that

$$w(\xi) = \frac{1}{\widetilde{e}(v)} \frac{|\xi|^{2v+d}}{\widetilde{K}_{v+d/2}(|\xi|)}, \tag{1.10}$$

$$c_\lambda e^{-b_\lambda |t|} \leq \widetilde{K}_\lambda(t) \leq c'_\lambda e^{-|t|}, \quad t \in \mathbb{R}^1. \tag{1.11}$$

Now if $\frac{1}{w} = \widehat{H}$ then

and w has property W1 for $\mathcal{A} = \{0\}$. Further, condition 1.7 holds iff $2\lambda - 2v - d > -d$ for $0 \leq \lambda \leq \kappa$ i.e. iff

$$-d/2 < v < 0. \tag{1.12}$$

and so when v satisfies 1.12, w has property W2 for all $\kappa \geq 0$.

<u>The Gaussian</u>

The Gaussian function

$$H(x) = \exp\left(-|x|^2\right), \quad x \in \mathbb{R}^d, \tag{1.13}$$

has Fourier transform

$$\widehat{H}(\xi) = \frac{\sqrt{\pi}}{2} \exp\left(-\frac{|\xi|^2}{4}\right), \quad \xi \in \mathbb{R}^d,$$

and so if $\frac{1}{w} = \widehat{H}$

$$w(\xi) = \frac{2}{\sqrt{\pi}} \exp\left(\frac{|\xi|^2}{4}\right), \tag{1.14}$$

and w has property W1 for empty $\mathcal{A}$ and property W2 for all $\kappa \geq 0$.

<u>Sobolev splines</u>

If v satisfies $v > d/2$ the multivariate Sobolev spline

$$H(x) = \frac{1}{2^{v-1}\Gamma(v)} \widetilde{K}_v(|x|), \quad x \in \mathbb{R}^d, \tag{1.15}$$

has Fourier transform

$$\widehat{H}(\xi) = \frac{1}{\left(1 + |\xi|^2\right)^v}, \quad \xi \in \mathbb{R}^d,$$

and so if $\frac{1}{w} = \widehat{H}$

$$w(\xi) = \left(1 + |\xi|^2\right)^v, \quad \xi \in \mathbb{R}^d, \tag{1.16}$$

and w has property W1 for empty $\mathcal{A}$ and property W2 for $0 \leq \kappa < v - d/2$.

1.2.4 Products of weight functions

See Theorem 33 where the product of weight functions is discussed in relation to the convolution of basis functions.

1.2.5 The tensor product of weight functions

The next theorem gives some general conditions under which the tensor product of two weight functions is a weight function.

Theorem 4 *Suppose w_1 and w_2 are weight functions i.e. they have property W1.*

1. *Then the tensor product $w_1 \otimes w_2$ also has property W1.*
2. *The tensor product $w_1 \otimes w_2$ has property W2 for parameter κ iff both w_1 and w_2 have property W2 for parameter κ.*

Proof. Part 1. In this proof meas denotes the measure of a set. Now suppose $w_i : \mathbb{R}^{d_i} \to \mathbb{R}$. Then each w_i is associated with a weight function set, call it $\mathcal{A}_i$, which is closed and has measure zero. The first step will be to construct the set $\mathcal{A}$ for $w_1 \otimes w_2$ from the sets $\mathcal{A}_i$. The candidate is

$$\mathcal{A} = \left(\mathcal{A}_1 \times \mathbb{R}^{d_2}\right) \cup \left(\mathbb{R}^{d_1} \times \mathcal{A}_2\right).$$

We show $\mathcal{A}$ is closed and has measure zero. We use the continuous projection operators $p_i : \mathbb{R}^d \to \mathbb{R}^{d_i}$ defined by $p_1(x) = x'$ and $p_2(x) = x''$. Next, since $\mathcal{A} = \bigcup\limits_{i=1}^{2} p_i^{-1}(\mathcal{A}_i)$ and each $\mathcal{A}_i$ is closed, $\mathcal{A}$ is also closed since p_i is continuous. Finally it must be shown that $\operatorname{meas}\mathcal{A} = 0$. But since

$$\operatorname{meas}\mathcal{A} = \operatorname{meas}\left(\mathcal{A}_1 \times \mathbb{R}^{d_2} \cup \mathbb{R}^{d_1} \times \mathcal{A}_2\right) \le \operatorname{meas}\left(\mathcal{A}_1 \times \mathbb{R}^{d_2}\right) + \operatorname{meas}\left(\mathbb{R}^{d_1} \times \mathcal{A}_2\right),$$

it is sufficient to show that $\operatorname{meas}\left(\mathcal{A}_1 \times \mathbb{R}^{d_2}\right) = 0$. To do this we use the countable additivity property of the Lebesgue measure on $\mathbb{R}^d$. In fact

$$\begin{aligned}
\operatorname{meas}\left(\mathcal{A}_1 \times \mathbb{R}^{d_2}\right) &= \operatorname{meas}\bigcup_{n=1}^{\infty}\left(\mathcal{A}_1 \times \left(\overline{B(0;n)} - B(0;n-1)\right)\right)\\
&= \sum_{n=1}^{\infty}\operatorname{meas}\left(\mathcal{A}_1 \times \left(\overline{B(0;n)} - B(0;n-1)\right)\right)\\
&= 0.
\end{aligned}$$

Part 2. Suppose w_1 and w_2 have property W2 for parameter κ. We show $w_1 \otimes w_2$ also has property W2 for parameter κ. Recall that property W2 requires that

$$\int \frac{|x|^{2\lambda}}{w_i(x)}dx < \infty, \quad 0 \le \lambda \le \kappa;\ i = 1,2.$$

Now

$$\int \frac{|x|^{2\lambda}}{w(x)}dx = \int \frac{\left(|x'|^2 + |x''|^2\right)^{\lambda}}{w_1(x')\,w_2(x'')}dx, \tag{1.17}$$

and the inequality

$$|x|^{2\lambda} \le 2^{\lambda}\left(|x'|^{2\lambda} + |x''|^{2\lambda}\right), \tag{1.18}$$

allows us to write

$$\begin{aligned}
\int \frac{|x|^{2\lambda}}{w(x)}dx &\le \int \frac{2^{\lambda}\left(|x'|^{2\lambda} + |x''|^{2\lambda}\right)}{w_1(x')\,w_2(x'')}dx\\
&= 2^{\lambda}\left(\int \frac{|x'|^{2\lambda}}{w_1(x')\,w_2(x'')}dx\right) + 2^{\lambda}\left(\int \frac{|x''|^{2\lambda}}{w_1(x')\,w_2(x'')}dx\right)\\
&= 2^{\lambda}\left(\int \frac{|x'|^{2\lambda}\,dx}{w_1(x')}\int \frac{dx}{w_2(x'')}\right) + 2^{\lambda}\left(\int \frac{dx'}{w_1(x')}\int \frac{|x''|^{2\lambda}\,dx''}{w_2(x'')}\right)\\
&< \infty.
\end{aligned}$$

Now suppose $w_1 \otimes w_2$ has property W2 for parameter κ. Thus, by 1.17

$$\int \frac{\left(|x'|^2 + |x''|^2\right)^\lambda}{w_1(x')\, w_2(x'')} dx < \infty, \quad 0 \le \lambda \le \kappa,$$

so that $\int \frac{|x'|^{2\lambda}}{w_1(x')\, w_2(x'')} dx < \infty$ and $\int \frac{|x''|^{2\lambda}}{w_1(x'')\, w_2(x'')} dx < \infty$.

Hence $\int \frac{|x'|^{2\lambda}}{w_1(x')} dx' < \infty$ and $\int \frac{|x''|^{2\lambda}}{w_2(x'')} dx'' < \infty$ i.e. w_1 and w_2 both have property W2 for parameter κ. ■

The next result generalizes the previous theorem to an arbitrary number of weight functions.

Corollary 5 *Suppose the functions $\{w_i\}_{i=1}^n$ satisfy property W1 of a weight function.*

1. *Then $\bigotimes_{i=1}^n w_i$ has property W1.*

2. *$\bigotimes_{i=1}^n w_i$ has property W2 for parameter κ iff each w_i has property W2 for parameter κ.*

Proof. An easy proof by induction using Theorem 4. ■

1.2.6 The hat weight function and the weight function properties

The next theorem shows under what conditions the multivariate hat function is a basis function of order zero.

Theorem 6 *Using the multivariate hat function Λ define the function w_Λ by*

$$w_\Lambda(\xi) = \frac{1}{\widehat{\Lambda}(\xi)}. \tag{1.19}$$

Then w_Λ satisfies the weight function properties W1 and W2 for the parameter κ in the following manner:

1. *w_Λ satisfies property W1 for all κ.*

2. *w_Λ satisfies property W2 for parameter κ iff $\kappa < 1/2$.*

Proof. Part 1 By 1.6, w_Λ is the set $\mathcal{A}$ which is closed with measure zero.

Part 2 Part 2 of Corollary 5 tells us that we need only establish property W2 in one-dimension. Now W2 holds for κ iff

$$\int_{\mathbb{R}^1} \frac{x^{2\lambda}}{w_\Lambda(x)} = \int x^{2\lambda} \frac{\sin^2(x/2)}{(x/2)^2} dx < \infty, \quad 0 \le \lambda \le \kappa.$$

But these integrals will exist iff they exist near infinity iff $2 - 2\lambda > 1$ i.e. iff $\lambda < 1/2$. ■

1.2.7 The extended B-spline weight functions

We now introduce a new class of weight functions which satisfy property W2. We call this class the *extended B-spline weight functions* because in Subsection 1.4.4 it will be shown that they generate basis functions which are the derivatives of the B-splines.

Theorem 7 ***The extended B-spline weight functions*** *Suppose $\kappa \ge 0$ is an integer. For given integers $l, n \ge 0$ define the extended B-spline weight function w by*

$$w(x) = \prod_{i=1}^d \frac{x_i^{2n}}{\sin^{2l} x_i}, \qquad x = (x_1, \ldots, x_d) \in \mathbb{R}^d. \tag{1.20}$$

Define the closed set of measure zero $\mathcal{A}$ to be the union of hyperplanes

$$\mathcal{A}=\bigcup_{\alpha\in\mathbb{Z}^d}\bigcup_{i=1}^{d}\{x:x_i=\pi\alpha\}.$$

Then the function w is a weight function with property W1 w.r.t. $\mathcal{A}$. Also, the weight function w has property W2 for κ iff n and l satisfy

$$\kappa+1/2<n\leq l. \tag{1.21}$$

Proof. Clearly property W1 is satisfied for $\mathcal{A}$ so w is a weight function. Corollary 5 tells us that we need only establish the criterion 1.21 in one dimension. Now W2 holds for κ iff

$$\int_{\mathbb{R}^1}\frac{x^{2\lambda}}{w(x)}=\int x^{2\lambda}\frac{\sin^{2l}x}{x^{2n}}dx<\infty,\quad 0\leq\lambda\leq\kappa. \tag{1.22}$$

But these integrals will exist if and only if they exist near the origin and near infinity. They will exist near the origin iff $2\lambda+2l-2n>-1$ i.e. iff $\lambda>n-l-1/2$, and they will exist near infinity iff $2\lambda-2n<-1$ i.e. iff $\lambda<n-1/2$. Thus the integrals all exist iff $n-l-1/2<\lambda<n-1/2$ for $0\leq\lambda\leq\kappa$ iff $\kappa<n-1/2$ and $n<l+1/2$ iff $\kappa+1/2<n\leq l$. ∎

1.3 The data spaces X_w^0

We now introduce the zero order data spaces X_w^0 which will later be used to define the minimum norm interpolant. Here w denotes the weight function. In Chapters 2, 4 and 5, X_w^0 will be used to define several variational interpolation and smoothing problems. In a manner analogous to the Sobolev spaces, mappings between X_w^0 and L^2 are used to show that X_w^0 is a Hilbert space. Various smoothness and C^∞ density results are then established.

Definition 8 ***The zero order inner product space*** X_w^0
Suppose w is a weight function i.e. it only has property W1 of Definition 1. Then define

$$X_w^0=\left\{f\in S':\widehat{f}\in L^1_{loc}\ \text{and}\ \sqrt{w}\widehat{f}\in L^2\right\}, \tag{1.23}$$

and endow it with the norm and inner product

$$\|f\|_{w,0}^2=\int w\left|\widehat{f}\right|^2,\qquad (f,g)_{w,0}=\int w\widehat{f}\overline{\widehat{g}}.$$

That $\|\cdot\|_{w,0}$ is a norm is simple to prove. Suppose $\|f\|_{w,0}=0$. Then, since $w>0$ a.e., $\int w\left|\widehat{f}\right|^2=0$ implies that $\widehat{f}(x)=0$ a.e. and thus $f=0$ in the distribution sense.

The next result reassures us that the space X_w^0 is non-empty.

Theorem 9 *If w is a weight function w.r.t. the closed set $\mathcal{A}$ of measure zero then*

$$\left\{\overset{\vee}{u}:u\in C_0^\infty\ \text{and}\ \operatorname{supp}u\subset\mathbb{R}^d\backslash\mathcal{A}\right\}\subset X_w^0, \tag{1.24}$$

where the set on the left is non-empty.

Proof. Firstly, $\mathbb{R}^d\backslash\mathcal{A}$ is not empty since $\mathcal{A}=\mathbb{R}^d$ implies $\operatorname{meas}\mathcal{A}\neq 0$. Clearly
$f\in\left\{\overset{\vee}{u}:u\in C_0^\infty\ \text{and}\ \operatorname{supp}u\subset\mathbb{R}^d\backslash\mathcal{A}\right\}$ implies $f\in S\subset S'$ and $\widehat{f}\in C_0^\infty\subset L^1_{loc}$. Also, since $\widehat{f}$ has bounded support in $\mathbb{R}^d\backslash\mathcal{A}$ and w is positive and continuous on $\mathbb{R}^d\backslash\mathcal{A}$

$$\int\left|\sqrt{w}\widehat{f}\right|^2=\int w\left|\widehat{f}\right|^2<\infty.$$

∎

1.3.1 The completeness of X_w^0

In a manner analogous to Sobolev space theory, the completeness of X_w^0 will be established by constructing an isometric homeomorphism $\mathcal{I}: X_w^0 \to L^2$ and then making use of the completeness of L^2.

Definition 10 ***The linear mappings $\mathcal{I}$ and $\mathcal{J}$***

Suppose w is a weight function. Using the definition of the space X_w^0 we can define the linear mapping $\mathcal{I}: X_w^0 \to L^2$ by

$$\mathcal{I}f = \left(\sqrt{w}\widehat{f}\right)^{\vee}, \quad f \in X_w^0.$$

If, in addition, we assume property W2 then $\frac{\widehat{g}}{\sqrt{w}} \in L^1 \subset S'$ when $g \in L^2$, and we can then define the linear mapping $\mathcal{J}: L^2 \to S'$ by

$$\mathcal{J}g = \left(\frac{\widehat{g}}{\sqrt{w}}\right)^{\vee}, \quad g \in L^2.$$

The linear mappings $\mathcal{I}$ and $\mathcal{J}$ have the following properties:

Theorem 11 *Suppose w has property W1 of a weight function. Then:*

1. *$\mathcal{I}: X_w^0 \to L^2$ is an isometry.*
2. *$\mathcal{I}$ is one-to-one.*

   Now suppose that in addition to property W1, w has property W2. Then:
3. *$\mathcal{J}: L^2 \to X_w^0$ and is an isometry.*
4. *$\mathcal{J}$ is one-to-one.*
5. *$\mathcal{J} \circ \mathcal{I} = I$ on X_w^0, and $\mathcal{I} \circ \mathcal{J} = I$ on L^2.*
6. *The mapping $\mathcal{I}: X_w^0 \to L^2$ is an isometric homeomorphism with inverse $\mathcal{J}$.*
7. *The mappings $\mathcal{I}$ and $\mathcal{J}$ are adjoints.*

Proof. The proofs are straight forward and will be omitted. ■

The fact that X_w^0 is complete, and hence is a Hilbert space, is a simple consequence of the last theorem:

Corollary 12 *Suppose w is a weight function with property W2. Then X_w^0 is complete and hence a Hilbert space.*

Proof. By part 6 of the previous Theorem 11, $\mathcal{I}: X_w^0 \to L^2$ is an isometric homeomorphism. Hence X_w^0 is complete since L^2 is a Hilbert space. ■

1.3.2 The smoothness of functions in X_w^0

We begin with a lemma of basic L^1 Fourier transform results.

Lemma 13

1. *(Theorem 4.2 of Malliavin [10]) If $f \in S'$ and $\widehat{f} \in L^1$, then $f \in C_B^{(0)}$, $|f(x)| \to 0$ as $|x| \to \infty$, and*

$$f(x) = (2\pi)^{-d/2} \int e^{ix\xi} \widehat{f}(\xi)\, d\xi.$$

2. *(Corollary 2.12 of Petersen [11]) If $f \in L^1$ and for some integer $n > 1$, $|\cdot|^n f \in L^1$, then $\overset{\vee}{f} \in C_B^{(n)}$. Further, $x^\beta f \in L^1$ when $|\beta| \le n$ and*

$$D^\beta \left(\overset{\vee}{f}\right)(x) = (i)^{|\beta|} \left(x^\beta f\right)^{\vee}(x), \quad |\beta| \le n. \tag{1.25}$$

The next theorem presents our first smoothness result for functions in X_w^0.

Theorem 14 *Suppose that w is a weight function with property W2 for parameter κ. Then $X_w^0 \subset C_B^{(\lfloor\kappa\rfloor)}$.*

Proof. Since w has property W2, by Theorem 3, if $|\beta| \leq \kappa$

$$\int \frac{|\cdot|^{2|\beta|}}{w} = \int\limits_{|\cdot|\leq 1} \frac{|\cdot|^{2|\beta|}}{w} + \int\limits_{|\cdot|\geq 1} \frac{|\cdot|^{2|\beta|}}{w} \leq \int\limits_{|\cdot|\leq 1} \frac{1}{w} + \int\limits_{|\cdot|\geq 1} \frac{|\cdot|^{2\kappa}}{w} < \infty. \tag{1.26}$$

Hence when $f \in X_w^0$ the Cauchy-Schwartz inequality yields

$$\int \left|\xi^\beta \widehat{f}\right| \leq \int \frac{|\cdot|^{|\beta|}}{\sqrt{w}} \sqrt{w} \left|\widehat{f}\right| \leq \left(\int \frac{|\cdot|^{2|\beta|}}{w}\right)^{1/2} \|f\|_{w,0} < \infty,$$

i.e. $f \in X_w^0$ implies $\widehat{D^\beta f} \in L^1$ when $|\beta| \leq \kappa$ and hence $f \in C_B^{(\lfloor\kappa\rfloor)}$ by part 1 of Lemma 13. Further, again by part 1 of Lemma 13

$$\left|D^\beta f(x)\right| = (2\pi)^{-d/2} \left|\int e^{ix\xi} \widehat{D^\beta f}(\xi)\, d\xi\right| \leq (2\pi)^{-d/2} \left|\int \left|\widehat{D^\beta f}(\xi)\right| d\xi\right|$$

■

The next theorem derives some inverse Fourier transform formulas and inequalities for derivatives of functions in X_w^0.

Theorem 15 *Suppose that w is a weight function with property W2 for parameter κ. Then for $f \in X_w^0$ we have the inverse Fourier transform formulas*

$$D^\beta f(x) = (2\pi)^{-d/2} \int e^{ix\xi} \widehat{D^\beta f}(\xi)\, d\xi, \quad |\beta| \leq \kappa. \tag{1.27}$$

and the derivatives satisfy

$$\left|D^\beta f(x)\right| \leq (2\pi)^{-d/2} \left(\int \frac{|\cdot|^{2|\beta|}}{w}\right)^{1/2} \|f\|_{w,0}, \quad |\beta| \leq \kappa, \tag{1.28}$$

so that $X_w^0 \hookrightarrow C_B^{(\lfloor\kappa\rfloor)}$ when $C_B^{(\lfloor\kappa\rfloor)}$ is endowed with the usual supremum norm.

Proof. From the proof of the previous theorem we have: $f \in X_w^0$ implies $\widehat{D^\beta f} \in L^1$ when $|\beta| \leq \kappa$. The inverse Fourier transform formulas 1.27 are now simple consequences of part 1 of Lemma 13.

The bounds on the derivatives are derived from 1.27 using the Cauchy-Schwartz inequality:

$$\left|D^\beta f(x)\right| \leq (2\pi)^{-\frac{d}{2}} \int \left|\xi^\beta\right| \left|\widehat{f}(\xi)\right| d\xi \leq (2\pi)^{-\frac{d}{2}} \int \frac{|\cdot|^{|\beta|}}{\sqrt{w}} \sqrt{w} \left|\widehat{f}\right| \leq (2\pi)^{-\frac{d}{2}} \left(\int \frac{|\cdot|^{2|\beta|}}{w}\right)^{\frac{1}{2}} \|f\|_{w,0}.$$

■

Remark 16 *Inequality 1.28 for $\beta = 0$ implies that X_w^0 is a reproducing kernel Hilbert space and in Section 1.5 we will calculate the Riesz representers of the evaluation functionals $f \to D^\alpha f(x)$. We have thus established an important link between pointwise processes and Hilbert space theory.*

1.3.3 Examples

Shifted thin-plate splines

From Examples 1.2.3, w has property W2 for all $\kappa \geq 0$, and thus Theorem 15 implies $X_w^0 \subset C_B^\infty$ as sets.

The Gaussian

It was shown in Examples 1.2.3 that w has property W2 for all $\kappa \geq 0$, and so Theorem 14 implies $X_w^0 \subset C_B^\infty$ as sets.

Also, by expanding the Gaussian about the origin using a Taylor series we can show that as sets $X_w^0 \subset H^\infty = \bigcap\limits_{n=0}^{\infty} H^n$, where H^n is the Sobolev space of order n (Definition 80).

Sobolev splines

It was shown in Examples 1.2.3 that property W2 holds for $0 \leq \kappa < v - d/2$. Now by Remark 16, $X_w^0 \hookrightarrow C_B^{(\max\lfloor\kappa\rfloor)}$ and we see that $\max\lfloor\kappa\rfloor = \lfloor v - d/2\rfloor$ when $v - d/2$ is not an integer and $\max\lfloor\kappa\rfloor = \lfloor v - d/2 - 1\rfloor = \lfloor v - d/2\rfloor - 1$ when $v - d/2$ is an integer.

Also, from 1.16, $w(\xi) = \left(1 + |\xi|^2\right)^v$ and so 1.23 implies $X_w^0 = H^v$ where H^v is the Sobolev space of positive order v defined using the Fourier transform.

Extended B-spline weight functions

These splines are defined in Theorem 7. Since property W2 holds for all $0 \leq \kappa < n - 1/2$ it follows that $\max\lfloor\kappa\rfloor = n - 1$, and thus $X_w^0 \hookrightarrow C_B^{(n-1)}$ by Remark 16.

In Lemma 89 we will show that $X_w^0 \hookrightarrow H_s^n$. Here H_s^n is the 'Sobolev-like' space introduced in Definition 81 below and it consists of all L^2 functions such that $D^\alpha f \in L^2$ when each $\alpha_i \leq n$.

1.3.4 Some dense C^∞ subspaces of X_w^0

The results of this subsection will not be used later in this document.

Results to handle the weight function discontinuity set

In this section we prove some results needed when we want to use a partition of unity to handle the discontinuity or zero values of a weight function on a set of measure zero. Recall that weight function property W1 introduced in Definition 1 required that a weight function be continuous and positive outside a closed set $\mathcal{A}$ of measure zero. Light and Wayne assumed their basis functions were continuous and positive outside the origin i.e. $\mathcal{A} = \{0\}$. In order to allow hat functions to be basis functions we have had to allow the weight set to a closed, unbounded set of measure zero.

The discontinuity of the weight function needs to be taken into account when we prove the density of some C^∞ subspaces in L^2 and X_w^0, and when we prove the smoothness of basis functions generated by weight functions which have property W2.

In this document we will use the next lemma whenever we want to construct a partition of unity using points 'near' a weight function set. This lemma will be used with $\mathcal{F} = \mathcal{A}$, where $\mathcal{A}$ is the weight function set. Lemma 18 will be applied to a function of the 'near' points. Nearness can be measured using the concept of a neighborhood of a set. In fact, if $\mathcal{F} \subset \mathbb{R}^d$ and $\eta > 0$ then define $\mathcal{F}_\eta = \bigcup\limits_{x \in A} B(x;\eta)$. The set $\mathcal{F}_\eta$ is referred to as the *η-neighborhood* of $\mathcal{F}$.

Lemma 17 *Let $\mathcal{F}$ be a set of points in $\mathbb{R}^d$. Then for any $\eta > 0$ there exists a function f_η such that:*

1. $f_\eta \in C^\infty$,
2. $0 \leq f_\eta(x) \leq 1$,
3. $f_\eta(x) = 1$ *when* $x \in \mathcal{F}_\eta$,
4. $f_\eta(x) = 0$ *when* $x \notin \mathcal{F}_{3\eta}$.

Proof. Choose $\eta > 0$. There exists a mollifier $\omega \in C_0^\infty$ satisfying

$$\operatorname{supp}\omega \subset B(0;1); \quad \int \omega = 1; \quad \omega \geq 0.$$

Now define $\omega_\eta(x) = \eta^{-d}\omega(x/\eta)$ and let $\chi_{2\eta}$ be the characteristic function of the set $\mathcal{F}_{2\eta}$. Then it can be shown that the function

$$f_\eta(x) = \int\limits_{\mathbb{R}^d} \chi_{2\eta}(y)\,\omega_\eta(x-y)\,dy,$$

has the required properties. ■

The next theorem will require the following standard measure theory results which are stated without proof.

Lemma 18 *1. Suppose the set $\mathcal{A}$ is closed and has measure zero. If $\mathcal{A}_\varepsilon$ is the $\varepsilon-$neighborhood of the set $\mathcal{A}$ then for any open ball B, $\operatorname{meas}(B \cap \mathcal{A}_\varepsilon) \to 0$ as $\varepsilon \to 0$.*

2. Suppose the set $\mathcal{A}$ is closed and has measure zero. Then $f \in L^1$ implies, $\lim_{\varepsilon\to 0} \int_{\mathcal{A}_\varepsilon} |f| = 0$.

Some spaces of continuous functions which are dense in L^2

The next theorem is an extension of Theorem 2.5 of Light and Wayne [8]. The difference is that instead of the set $\{0\}$, we are dealing with a set $\mathcal{A}$ which is closed and has measure zero. There is no need to assume that $0 \in \mathcal{A}$. To handle this set we use Lemma 17 and Lemma 18.

Theorem 19 *Suppose w is a weight function with properties W1 and W2 with respect to weight set $\mathcal{A}$. Then the set $\left\{\sqrt{w}f : f \in C_0^{(0)}\right\} \cap L^2$ is dense in L^2.*

Here $C_0^{(0)}$ denotes the continuous functions with compact support.

Proof. In this proof we will use the abbreviation

$$\sqrt{w}C_0^{(0)} \cap L^2 = \left\{\sqrt{w}f : f \in C_0^{(0)}\right\} \cap L^2.$$

By Lemma 17, given $h > 0$, there exists a function $\psi_h \in C^\infty$ satisfying $0 \le \psi_h \le 1$, $\psi_h = 1$ on $\mathcal{A}_h$ and $\psi_h = 0$ outside $\mathcal{A}_{3h}$. Now define the mapping $\Psi_h : C_0^{(0)} \to C_0^{(0)}$ by

$$\Psi_h g = (1 - \psi_h)\, g, \quad g \in C_0^{(0)},\ h > 0.$$

Since w is positive and continuous on $\mathbb{R}^d \backslash \mathcal{A}$, we have $\frac{(1-\psi_h)}{\sqrt{w}} \in C^{(0)}$ and hence $(1 - \psi_h)\, g \in \sqrt{w}C_0^{(0)} \cap L^2$. In other words, $\Psi_h : C_0^{(0)} \to \sqrt{w}C_0^{(0)} \cap L^2$.

Now $C_0^{(0)}$ is dense in L^2, and given $g_0 \in L^2$ and $\varepsilon > 0$ we can choose $g_\varepsilon \in C_0^{(0)}$ such that $\|g_0 - g_\varepsilon\|_{L^2} < \varepsilon/2$. Then

$$\|g_0 - \Psi_h g_\varepsilon\|_{L^2} \le \|g_0 - g_\varepsilon\|_{L^2} + \|g_\varepsilon - \Psi_h g_\varepsilon\|_{L^2} \le \varepsilon/2 + \|\psi_h g_\varepsilon\|_{L^2} \le \varepsilon/2 + \left(\int_{\mathcal{A}_{3h}} |g_\varepsilon|^2 \right)^{\frac{1}{2}}.$$

Finally, since $|g_\varepsilon|^2 \in L^1$, it follows from Lemma 18 that $\lim_{h\to 0} \int_{\mathcal{A}_{3h}} |g_\varepsilon|^2 = 0$. Thus we can choose $h > 0$ so that $\|g_0 - \Psi_h g_\varepsilon\|_{L^2} \le \varepsilon$. ■

The next theorem improves the previous theorem. This is Light and Wayne's Proposition 2.7 and our proof is a variation of Light and Wayne's proof.

Theorem 20 *Suppose w has extended weight function properties W1 and W2 with respect to weight set $\mathcal{A}$. Then the set $\{\sqrt{w}f : f \in C_0^\infty\} \cap L^2$ is dense in L^2.*

Proof. In this proof we will use the abbreviations,
$\sqrt{w}C_0^\infty \cap L^2 = \{\sqrt{w}f : f \in C_0^\infty\} \cap L^2$ and $\sqrt{w}C_0^{(0)} \cap L^2 = \left\{\sqrt{w}f : f \in C_0^{(0)}\right\} \cap L^2$.

Now $\sqrt{w}C_0^\infty \cap L^2 \subset \sqrt{w}C_0^{(0)} \cap L^2$. Hence, by Theorem 19, if it can be shown that $\sqrt{w}C_0^\infty \cap L^2$ is dense in $\sqrt{w}C_0^{(0)} \cap L^2$, this theorem is true. First note that C_0^∞ is dense in L^2 and hence dense in $C_0^{(0)}$. Next select $\varepsilon > 0$ and define a mapping $\Theta_\varepsilon : C_0^{(0)} \to C_0^\infty$ where $\Theta_\varepsilon f$ is an element of C_0^∞ such that $\|f - \Theta_\varepsilon f\|_{L^2} < \varepsilon$.

We want to use Θ_ε to construct a mapping (see left side of 1.29) from $\sqrt{w}C_0^{(0)} \cap L^2$ to $\sqrt{w}C_0^\infty \cap L^2$, which can be used to prove that $\sqrt{w}C_0^\infty \cap L^2$ is dense in $\sqrt{w}C_0^{(0)} \cap L^2$. If $f \in \sqrt{w}C_0^{(0)} \cap L^2$, then $\frac{f}{\sqrt{w}} \in C_0^{(0)}$ and $\sqrt{w}\Theta_\delta\left(\frac{f}{\sqrt{w}}\right) \in \sqrt{w}C_0^\infty$, $\delta > 0$. But we still need $\sqrt{w}\Theta_\delta\left(\frac{f}{\sqrt{w}}\right) \in L^2$. To do this we use the function ψ_h defined in the previous Theorem 19. Since $(1 - \psi_h)\sqrt{w} \in C^{(0)}$ we have

$$(1 - \psi_h)\sqrt{w}\Theta_\delta\left(\frac{f}{\sqrt{w}}\right) \in \sqrt{w}C_0^\infty \cap L^2,\ \text{when } f \in \sqrt{w}C_0^{(0)} \cap L^2.$$

Finally, it will turn out that the support of $(1-\psi_h)\sqrt{w}\Theta_\delta\left(\frac{f}{\sqrt{w}}\right)$ needs to be restricted by multiplying it by a function $\phi_R \in C_0^\infty$, $R \geq 1$, such that $0 \leq \phi_R \leq 1$, $\operatorname{supp}\phi_R \subset [-R-1, R+1]$ and $\phi_R = 1$ on $[-R, R]$.

We now assert that there exist R, h and δ such that $\left\|(1-\psi_h)\phi_R\sqrt{w}\Theta_\delta\left(\frac{f}{\sqrt{w}}\right) - f\right\|_{L^2} < \varepsilon$. To prove this assertion write

$$\begin{aligned}(1-\psi_h)\phi_R\sqrt{w}\Theta_\delta\left(\frac{f}{\sqrt{w}}\right) - f &= (1-\psi_h)\phi_R\sqrt{w}\Theta_\delta\left(\frac{f}{\sqrt{w}}\right) - (1-\psi_h)\phi_R\sqrt{w}\left(\frac{f}{\sqrt{w}}\right) + \\ &\quad + (1-\psi_h)\phi_R\sqrt{w}\left(\frac{f}{\sqrt{w}}\right) - \phi_R f + \phi_R f - f \\ &= (1-\psi_h)\phi_R\sqrt{w}\left(\Theta_\delta\left(\frac{f}{\sqrt{w}}\right) - \frac{f}{\sqrt{w}}\right) + \psi_h\phi_R f + (\phi_R - 1) f.\end{aligned}$$

Observe that $(1-\psi_h)\sqrt{w} \in C^{(0)}$ implies $(1-\psi_h)\phi_R\sqrt{w} \in C_0^{(0)}$. Hence

$$\begin{aligned}\left\|(1-\psi_h)\phi_R\sqrt{w}\Theta_\delta\left(\frac{f}{\sqrt{w}}\right) - f\right\|_{L^2} &\leq \left\|(1-\psi_h)\phi_R\sqrt{w}\left(\Theta_\delta\left(\frac{f}{\sqrt{w}}\right) - \frac{f}{\sqrt{w}}\right)\right\|_{L^2} + \\ &\quad + \|\psi_h f\|_{L^2} + \|(\phi_R - 1) f\|_{L^2} \\ &\leq \left\|(1-\psi_h)\phi_R\sqrt{w}\right\|_\infty \left\|\Theta_\delta\left(\frac{f}{\sqrt{w}}\right) - \frac{f}{\sqrt{w}}\right\|_{L^2} + \\ &\quad + \|\psi_h f\|_{L^2} + \|(\phi_R - 1) f\|_{L^2} \\ &\leq \left\|(1-\psi_h)\phi_R\sqrt{w}\right\|_\infty \delta + \|\psi_h f\|_{L^2} + \|(\phi_R - 1) f\|_{L^2}.\end{aligned}$$

We will now consider each term on the right side of the last inequality separately. Regarding the last term, $\|(\phi_R - 1) f\|_{L^2} = \left(\int_{|\cdot|\geq R} |f|^2\right)^{1/2}$ and so R can be fixed so that $\|(\phi_R - 1) f\|_{L^2} \leq \varepsilon/3$. Next, by definition of ψ_h, $\|\psi_h f\|_{L^2} = \left(\int_{\mathcal{A}_{3h}} |f|^2\right)^{1/2}$ and, since $|f|^2 \in L^1$, it follows from Lemma 18 that $\lim\limits_{h\to 0} \int_{\mathcal{A}_{3h}} |f|^2 = 0$. A value of h can now be chosen so that $\|\psi_h f\|_{L^2} \leq \varepsilon/3$. Now we have

$$\left\|(1-\psi_h)\phi_R\sqrt{w}\Theta_\delta\left(\frac{f}{\sqrt{w}}\right) - f\right\|_{L^2} \leq \left\|(1-\psi_h)\phi_R\sqrt{w}\right\|_\infty \delta + 2\varepsilon/3, \tag{1.29}$$

for all $f \in \sqrt{w}C_0^{(0)} \cap L^2$.

The last step is to choose δ so that $\|(1-\psi_h)\phi_R\sqrt{w}\|_\infty \delta \leq \varepsilon/3$, and the theorem follows. ■

1.3.5 Some dense C_0^∞ and S subspaces of X_w^0

The next result corresponds to Light and Wayne's Corollary 2.8. Here X_w^0 is Light and Wayne's space Y and $\widehat{C_0^\infty}$ denotes the space of Fourier transforms of all functions in C_0^∞.

Corollary 21 *Suppose w has properties W1 and W2 i.e. $w \in C^{(0)}\left(\mathbb{R}^d \backslash \mathcal{A}\right)$ and $w > 0$ on $\mathbb{R}^d \backslash \mathcal{A}$, where $\mathcal{A}$ is a closed set of measure zero.*

Then the spaces $X_w^0 \cap \widehat{C_0^\infty}$ and $X_w^0 \cap S$ are dense X_w^0.

Proof. Let $\mathcal{I} : X_w^0 \to L^2$ be as given in Definition 10 i.e. $\mathcal{I}f = \sqrt{w}\widehat{f}$. Then

$$\mathcal{I}\left(X_w^0 \cap \widehat{C_0^\infty}\right) = \left\{\sqrt{w}\widehat{f} : f \in X_w^0 \text{ and } \widehat{f} \in C_0^\infty\right\} = \left\{\sqrt{w}\widehat{f} : \widehat{f} \in C_0^\infty\right\} \cap L^2.$$

Thus, by parts 1 and 2 of Theorem 11, $\mathcal{I}$ is an isometric isomorphism from $X_w^0 \cap \widehat{C_0^\infty}$ to $\left\{\sqrt{w}\widehat{f} : \widehat{f} \in C_0^\infty\right\} \cap L^2$. Since we know by Theorem 20 that the last set is dense in L^2, we have the density of $X_w^0 \cap \widehat{C_0^\infty}$ in X_w^0.

Clearly, since $X_w^0 \cap \widehat{C_0^\infty} \subset X_w^0 \cap S$, $X_w^0 \cap S$ is dense in $\mathbb{X}_w^0$. ■

1.4 Basis functions

In this section we will define a (unique) basis function of order zero generated by a weight function with property W2, and note that this contrasts with the positive order case where the basis function is only unique modulo a space of polynomials. Basis functions are generally denoted by G and will always be members of X_w^0. If a weight function has property W2 for parameter κ it will be shown that $G \in C_B^{(\lfloor 2\kappa \rfloor)}$, whereas in general we know that $X_w^0 \subset C_B^{(\lfloor \kappa \rfloor)}$. Basis functions are used to construct solutions to the variational interpolation problem of part 2 of this document and to the two smoothing problems studied subsequently in Chapters 4 and 5.

We will also derive results concerning the basis functions generated by the tensor (or direct) product of weight functions with reference to the example of the extended B-spline basis functions introduced in Subsection 1.2.7.

Definition 22 ***Basis function***

Suppose a weight function w has property W2. Then $\frac{1}{w} \in L^1$ and hence by Lemma 13, $\left(\frac{1}{w}\right)^{\vee} \in C_B^{(0)}$. We now define the unique basis function G of order 0 generated by w to be

$$G = \left(\frac{1}{w}\right)^{\vee}. \tag{1.30}$$

A simple consequence is that $\overline{G(x)} = G(-x)$.

The next theorem makes three important points: (1) unlike in the positive order case basis functions are unique and are data functions i.e. they are members of X_w^0, (2) scaling can be used to enlarge the classes of weight functions available and (3) in the case of the extended B-splines scaling can dramatically improve the performance of the interpolant.

Theorem 23 *Suppose w has property W2 for κ. Then:*

1. *The basis function G generated by the weight function w is a member of X_w^0.*
2. *If $\lambda > 0$ then the weight function $\lambda^d w\left(\frac{t}{\lambda}\right)$ has property W2 for κ iff w has property W2 for κ, and $G(\lambda x)$ is the basis function generated by $\lambda^d w\left(\frac{t}{\lambda}\right)$.*

Proof. Since $C_B^{(0)} \subset S'$ it follows that $G \in S'$, $\widehat{G} \in L_{loc}^1$ and $\sqrt{w}\widehat{G} = \frac{1}{\sqrt{w}} \in L^2$ i.e. $G \in X_w^0$.

Further, $\widehat{G(\lambda t)} = \frac{1}{\lambda^d}\widehat{G}\left(\frac{t}{\lambda}\right) = \frac{1}{\lambda^d w\left(\frac{t}{\lambda}\right)}$ so the corresponding weight function is $\lambda^d w\left(\frac{t}{\lambda}\right)$. But by part 6 of Remark 2 this is a weight function with the same parameter κ as w. ■

1.4.1 The smoothness of basis functions

Not only is the basis function a data function but all basis function derivatives up to order κ are data functions:

Theorem 24 *Suppose the weight function w has property W2 for parameter κ. Then the basis function G generated by w is in $C_B^{(\lfloor 2\kappa \rfloor)}\left(\mathbb{R}^d\right)$. Additionally, the inverse Fourier transform formulas*

$$D^{\beta}G(x) = \frac{1}{(2\pi)^{d/2}} \int e^{ix\xi} \widehat{D^{\beta}G}(\xi)\, d\xi, \quad |\beta| \le 2\kappa, \tag{1.31}$$

hold and $D^{\beta}G \in X_w^0$ when $|\beta| \le \kappa$.

Proof. Since property W2 implies $\frac{1}{w} \in L^1$ and $\frac{|\cdot|^{2\kappa}}{w} \in L^1$, an application of the formula of part 2 of Lemma 13 with $f = \frac{1}{w}$ proves $G \in C_B^{(\lfloor 2\kappa \rfloor)}$. Formula 1.31 then follows from equation 1.25 of Lemma 13.

Now to prove $D^{\beta}G \in X_w^0$. Firstly, $G \in S'$ implies $D^{\beta}G \in S'$. Next, for compact K, $\int_K \left|\widehat{D^{\beta}G}\right| \le \int_K \frac{|\cdot|^{|\beta|}}{w} < \infty$ since $\frac{1}{w} \in L^1$. Finally, for $|\beta| \le \kappa$, 1.26 implies

$$\left\|D^{\beta}G\right\|_{w,0}^2 = \int w \left|\widehat{D^{\beta}G}\right|^2 = \int \frac{\xi^{2\beta}}{w} \le \int \frac{|\cdot|^{2|\beta|}}{w} < \infty, \tag{1.32}$$

since w has property W2. ■

Remark 25

1. *If the set of values of* 2κ *has a finite upper bound then although* $\max 2\kappa$ *may not exist e.g. the extended B-splines example below,* $\max \lfloor 2\kappa \rfloor$ *always exists and* $G \in C_B^{(\max\lfloor 2\kappa \rfloor)}$.
2. *Allowing* κ *to take non-integer values in the definition of weight property W2 (Definition 1) can yield an extra order of smoothness e.g. the Sobolev spline example below.*

1.4.2 Examples: radial basis functions

The radial functions used in Examples 1.2.3 to generate weight functions (generally denoted by G), namely the shifted thin-plate splines, Gaussian and Sobolev splines (univariate Laplacian kernels) can be used as radial basis functions. The basis functions associated with the extended B-spline weight functions will be derived below in Subsection 1.4.4.

Shifted thin-plate splines

The basis functions are given by 1.8 where $-d/2 < v < 0$ i.e.

$$G(x) = (-1)^{\lceil v \rceil} \left(1 + |x|^2\right)^v, \quad -d/2 < v < 0. \tag{1.33}$$

The weight function has property W2 for all $\kappa \geq 0$. Hence by Theorem 24, the basis function lies in $C_B^{(\lfloor 2\kappa \rfloor)}$ for $\kappa \geq 0$ and so is a C_B^∞ function.

The Gaussian

The weight function 1.14 has property W2 for $\kappa \geq 0$ and the basis function is given by 1.13 i.e. $G(x) = e^{-|x|^2}$. Hence by Theorem 24, we have $G \in C_B^{(\lfloor 2\kappa \rfloor)}$ for $\kappa \geq 0$ and so $G \in C_B^\infty$. In fact $G \in S$.

Sobolev splines

The basis function is 1.15 i.e.

$$G(x) = \frac{1}{2^{v-1}\Gamma(v)} \widetilde{K}_{v-d/2}(|x|), \quad x \in \mathbb{R}^d,$$

where $v > d/2$. It was shown in Examples 1.2.3 that property W2 holds for all $0 \leq \kappa < v - d/2$. Now by Remark 25, $G \in C_B^{(\max\lfloor 2\kappa \rfloor)}$ and we see that $\max \lfloor 2\kappa \rfloor = \lfloor 2v - d \rfloor$ when $2v$ is not an integer and $\max \lfloor 2\kappa \rfloor = \lfloor 2v - d \rfloor - 1$ when $2v$ is an integer. This is an example where allowing κ to take non-integer values in the definition of weight property W2 (Definition 1) yields an extra order of smoothness. An example is when $v - d/2 = 1.6$.

1.4.3 Basis functions generated by tensor product weight functions

Here we prove that the basis function of a tensor product weight function is the tensor product of basis functions.

Theorem 26 *Suppose* w_1 *and* w_2 *are weight functions which satisfy property W2 for parameter* κ*. Then part 2 of Theorem 4 implies that* $w = w_1 \otimes w_2$ *is a tensor product weight function satisfying property W2 for parameter* κ*.*

We prove here that $G = G_1 \otimes G_2$ *where* G_1, G_2 *and* G *are the basis functions of order* 0 *generated by* w_1, w_2 *and* w *respectively.*

Proof. Here each $G_i \in C_B^{(0)}$ and so each $G_i \in S'$ and we are interested in the tensor product of members of S'. In this proof we refer to results from Vladimirov [12] where the term *direct products* is used instead of *tensor products* and S' is referred to as the space of generalized functions of slow growth. In Subsection 2.8.5 of Vladimirov it is shown that $G_1 \otimes G_2 \in S'$ and in part (e) of Subsection 2.9.3 it is shown that $\widehat{G} = \widehat{G_1} \otimes \widehat{G_2}$. Thus

$$\widehat{G} = \widehat{G_1} \otimes \widehat{G_2} = \frac{1}{w_1} \otimes \frac{1}{w_2} = \frac{1}{w_1 \otimes w_2} = \frac{1}{w}.$$

■

Corollary 27 *Suppose that $\{w_i\}$ is a set of weight functions which satisfy W2 for parameter κ. Then by part 2 of Theorem 4 $w = \otimes w_i$ is a tensor product weight function satisfying property W2 for parameter κ.*

Further suppose that $\{G_i\}$ is the set of basis functions of order 0 generated by the w_i.

We claim here that $G = \otimes G_i$ is a basis function of order zero generated by w.

Proof. By induction using Theorem 26. ■

One of the main motivations of this document is to define the weight functions so that tensor product hat functions are basis functions. The following corollary shows that the hat function is a basis function in any dimension.

Corollary 28 *Suppose Λ is the $d-$dimensional tensor product hat function. Then Λ is the basis function of order zero generated by the weight function $\frac{1}{\Lambda}$.*

1.4.4 The extended B-spline basis functions

In this subsection we will derive the basis functions of zero order generated by the class of tensor product weight functions studied in Subsection 1.2.7, namely the extended B-spline weight functions. The extended B-spline weight functions were so named because their basis functions will turn out to be the derivatives of the convolutions of hat functions, denoted $(*\Lambda)^l$, $\quad l = 1, 2, 3, \ldots$ and these are the B-splines. The next theorem gives the unique tensor product basis function of zero order generated by these weight functions.

Theorem 29 ***The extended B-spline basis functions*** *Suppose w is the extended B-spline weight function considered in Theorem 7 and suppose w has property W2 for κ i.e. $\kappa + 1/2 < n \le l$.*

Then the basis function G of order zero generated by w is the tensor product $G(x) = \prod\limits_{k=1}^{d} G_1(x_k)$ where

$$G_1(t) = (-1)^{l-n} \tfrac{(2\pi)^{l/2}}{2^{2(l-n)+1}} D^{2(l-n)} \left((*\Lambda)^l \right) \left(\tfrac{t}{2} \right), \quad t \in \mathbb{R}^1, \tag{1.34}$$

and $(*\Lambda)^l$ denotes the convolution of l 1-dimensional hat functions. Further, if $n < l$ we have

$$D^{2(l-n)} (*\Lambda)^l = \frac{(-1)^{l-n}}{(2\pi)^{(l-n)/2}} (*\Lambda)^{n-1} * \sum_{k=-(l-n)}^{l-n} (-1)^k \tbinom{2(l-n)}{l-n+k} \Lambda(\cdot - k),$$

$G_1 \in C_0^{(2n-2)}(\mathbb{R}^1)$, $D^{2n-1} G_1$ is a bounded, piecewise constant function with bounded support, and $D^{2n} G_1$ is the finite sum of translated delta functions.

Finally, $G \in C_0^{(2n-2)}(\mathbb{R}^d)$, the $\{D^\alpha G\}_{|\alpha|=2n-1}$ are bounded functions with bounded support, and the$\{D^\alpha G\}_{|\alpha|=2n}$ are the tensor product of bounded functions with bounded support, and finite sums of translated delta functions.

Proof. In this proof it will be better to use the operator notation $F[f]$ for the Fourier transform instead of $\widehat{f}$.

Define $w_1(t) = \frac{t^{2n}}{\sin^{2l} t}$, $\quad t \in \mathbb{R}$. Since the 1-dimensional hat function Λ satisfies

$$F[\Lambda](t) = (2\pi)^{-1/2} \left(\frac{\sin(t/2)}{t/2} \right)^2, \qquad t \in \mathbb{R},$$

we have

$$\begin{aligned}
\tfrac{1}{w_1(t)} = \tfrac{\sin^{2l} t}{t^{2n}} = t^{2(l-n)} \left(\tfrac{\sin t}{t} \right)^{2l} &= (2\pi)^{l/2} t^{2(l-n)} (F[\Lambda](2t))^l \\
&= (2\pi)^{l/2} t^{2(l-n)} F\left[(*\Lambda)^l \right](2t) \\
&= (2\pi)^{l/2} \left(\tfrac{1}{2} \right)^{2(l-n)} \left(t^{2(l-n)} F\left[(*\Lambda)^l \right] \right)(2t) \\
&= (2\pi)^{l/2} \left(\tfrac{i}{2} \right)^{2(l-n)} F\left[D^{2(l-n)} (*\Lambda)^l \right](2t) \\
&= (-1)^{l-n} \tfrac{(2\pi)^{l/2}}{2^{2(l-n)}} F\left[D^{2(l-n)} (*\Lambda)^l \right](2t),
\end{aligned}$$

Now since the parameters n and l used to define the weight function w are independent of the dimension d, w_1 satisfies properties W1 to W2 and so $1/w_1 \in L^1$. Also

$$\begin{aligned}G_1(t) = F^{-1}\left[\frac{1}{w_1}\right](t) &= (-1)^{l-n}\,\tfrac{(2\pi)^{l/2}}{2^{2(l-n)}}F^{-1}\left[F\left[D^{2(l-n)}(*\Lambda)^l\right](2t)\right](t)\\ &= (-1)^{l-n}\,\tfrac{(2\pi)^{l/2}}{2^{2(l-n)}}\frac{1}{2}D^{2(l-n)}(*\Lambda)^l\left(\frac{t}{2}\right)\\ &= (-1)^{l-n}\,\tfrac{(2\pi)^{l/2}}{2^{2(l-n)+1}}D^{2(l-n)}(*\Lambda)^l\left(\frac{t}{2}\right).\end{aligned}$$

We will require the convolution identities

$$\delta(\cdot - a) * f = (2\pi)^{-1/2} f(\cdot - a), \quad f \in \mathcal{D}',$$

and

$$(*(\delta(\cdot - 1) - 2\delta + \delta(\cdot + 1)))^m = \frac{(-1)^m}{(2\pi)^{(m-1)/2}} \sum_{k=-m}^{m} (-1)^k \tbinom{2m}{m+k}\delta(\cdot - k).$$

Now suppose that $n < l$. Then in one dimension

$$\begin{aligned}D^{2(l-n)}(*\Lambda)^l = (*\Lambda)^n * \left(*D^2\Lambda\right)^{l-n} &= (*\Lambda)^n * (*(\delta(\cdot - 1) - 2\delta + \delta(\cdot + 1)))^{l-n}\\ &= \frac{(-1)^{l-n}}{(2\pi)^{(l-n-1)/2}}(*\Lambda)^n * \sum_{k=-(l-n)}^{l-n}(-1)^k\tbinom{2(l-n)}{l-n+k}\delta(\cdot - k)\\ &= \frac{(-1)^{l-n}}{(2\pi)^{(l-n)/2}}(*\Lambda)^{n-1} * \sum_{k=-(l-n)}^{l-n}(-1)^k\tbinom{2(l-n)}{l-n+k}\Lambda(\cdot - k).\end{aligned}$$

Further

$$\begin{aligned}D^{2n-2}G_1(t) &= (-1)^{l-n}\,\tfrac{(2\pi)^{l/2}}{2^{2l+1}}\left(D^{2n-2}D^{2(l-n)}(*\Lambda)^l\right)\left(\frac{t}{2}\right)\\ &= (-1)^{l-n}\,\tfrac{(2\pi)^{l/2}}{2^{2l+1}}\left(D^{2l-2}(*\Lambda)^l\right)\left(\frac{t}{2}\right)\\ &= (-1)^{l-n}\,\tfrac{(2\pi)^{l/2}}{2^{2l+1}}\left(\left(*D^2\Lambda\right)^{l-1} * \Lambda\right)\left(\frac{t}{2}\right)\\ &= (-1)^{l-n}\,\tfrac{(2\pi)^{l/2}}{2^{2l+1}}\left((*(\delta(\cdot - 1) - 2\delta + \delta(\cdot + 1)))^{l-1} * \Lambda\right)\left(\frac{t}{2}\right)\\ &= (-1)^{n-1}\,\tfrac{(2\pi)^{l/2}}{2^{2l+1}}\frac{(-1)^{l-1}}{(2\pi)^{\frac{l-2}{2}}}\sum_{\substack{k=\\-(l-1)}}^{l-1}(-1)^k\tbinom{2(l-1)}{l-1+k}(\delta(\cdot - k) * \Lambda)\left(\frac{t}{2}\right)\\ &= (-1)^{n-1}\,\tfrac{(2\pi)^{l/2}}{2^{2l+1}}\frac{(-1)^{l-1}}{(2\pi)^{\frac{l-1}{2}}}\sum_{k=-(l-1)}^{l-1}(-1)^k\tbinom{2(l-1)}{l-1+k}\Lambda\left(\frac{t}{2} - k\right)\\ &= (-1)^{n+l}\,\tfrac{(2\pi)^{1/2}}{2^{2l+1}}\sum_{k=-(l-1)}^{l-1}(-1)^k\tbinom{2(l-1)}{l-1+k}\Lambda\left(\frac{t}{2} - k\right),\end{aligned}$$

which is a continuous, piecewise linear function with bounded support. The stated properties of G_1 and G now follow directly. ■

1.4.5 *Convolutions and products of weight and basis functions*

The results of this subsection will not be used later in this document. In this subsection we give several results concerning the product and convolution of basis functions. The next result gives conditions under which the product of two basis functions of order zero is a basis function of order zero.

Theorem 30 *Suppose the weight functions w_1 and w_2 have property W2 for parameter κ_i. Suppose G_1 and G_2 are the basis functions of order zero generated by the weight functions w_1 and w_2. Suppose that $1/w_1$ is bounded a.e. or $G_1 G_2 \in L^1$.*

Then the product $G_1 G_2$ is a basis function of order zero generated by a weight function w which satisfies property W2 for parameter $\kappa = \min\{\kappa_1, \kappa_2\}$.

Further, $w \in C^{(0)}$ and $w(x) > 0$ for all x i.e. $\widehat{G_1 G_2}(\xi) > 0$ for all ξ.

Proof. We first note that from Theorem 2.2 and Exercise 2.4 of Petersen [11], that if $f, g \in L^1$ then $f * g \in L^1$, $\widehat{f * g} = \widehat{f}\widehat{g}$ and, if f is bounded a.e., $f * g$ is a continuous, bounded function.

Now define the functions G and w by $\frac{1}{w} = \frac{1}{w_1} * \frac{1}{w_2} \in L^1$ and $\widehat{G} = \frac{1}{w}$. We now have $\left(\frac{1}{w}\right)^{\vee} = \left(\frac{1}{w_1}\right)^{\vee} \left(\frac{1}{w_2}\right)^{\vee}$ i.e. $G = G_1 G_2$.

If $G_1 G_2 \in L^1$ then $\frac{1}{w} \in C_B^{(0)}$, and if $\frac{1}{w_1}$ is bounded a.e. then $\frac{1}{w_1} * \frac{1}{w_2} = \frac{1}{w} \in C_B^{(0)}$. Thus $|w(x)| \geq c > 0$ for some constant c, and is continuous whenever w is finite. But $w_1 \geq 0$ and $w_2 \geq 0$ a.e. so $w(x) \geq c$ for all x.

The next step is to show that $w(x) < \infty$ for all x. By definition of w_1 and w_2 there exist closed sets of measure zero, $\mathcal{A}_1$ and $\mathcal{A}_2$ such that w_i is continuous and positive outside $\mathcal{A}_i$. For each $x \in \mathbb{R}^d$, let $\mathcal{B}_x = (x - \mathcal{A}_1) \cup \mathcal{A}_2$ and note that $\mathcal{B}_x$ is closed with measure zero, and that, as a function of y, $w_1(x-y)\, w_2(y)$ is continuous and positive on $\mathbb{R}^d \backslash \mathcal{B}_x$. Thus, we have for all x

$$\frac{1}{w(x)} = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d \backslash \mathcal{B}_x} \frac{dy}{w_1(x-y)\, w_2(y)} > 0, \tag{1.35}$$

and consequently that $w(x) < \infty$ for all x i.e. $w \in C^{(0)}(\mathbb{R})$.

It remains to be shown that w satisfies property W2 for $\kappa = \min\{\kappa_1, \kappa_2\}$. Suppose $0 \leq \lambda \leq \kappa$. Then, for some constant $c_\lambda > 0$

$$\begin{aligned}
\int \frac{|x|^{2\lambda}\, dx}{w(x)} &\leq c_\lambda \int\int \frac{|x-y|^{2\lambda}\, dy\, dx}{w_1(x-y)\, w_2(y)} + c_\lambda \int\int \frac{|y|^{2\lambda}\, dy\, dx}{w_1(x-y)\, w_2(y)} \\
&= c_\lambda \int \frac{|x-y|^{2\lambda}\, dx}{w_1(x-y)} \int \frac{dy}{w_2(y)} + c_\lambda \int \frac{|y|^{2\lambda}\, dy}{w_2(y)} \int \frac{dx}{w_1(x-y)} \\
&= c_\lambda \int \frac{|\cdot|^{2\lambda}}{w_1} \int \frac{1}{w_2} + c_\lambda \int \frac{|\cdot|^{2\lambda}}{w_2} \int \frac{1}{w_1} \\
&< \infty.
\end{aligned}$$

■

The last theorem can be expressed as the following weight function result.

Corollary 31 *Suppose w_1 and w_2 are two weight functions which satisfy W2 for $\kappa = \kappa_1$ and $\kappa = \kappa_2$ respectively. Further, suppose that $1/w_1$ is bounded. Define the function w by $\frac{1}{w} = \frac{1}{w_1} * \frac{1}{w_2}$.*

Then w is a weight function which has property W2 for $\kappa = \min\{\kappa_1, \kappa_2\}$.

Further, $w \in C^{(0)}(\mathbb{R})$ and $w(x) > 0$ for all x.

Remark 32 *Any basis function can be uniformly pointwise approximated by a sequence of basis functions which have Fourier transforms that are always positive.*

In fact, suppose G is a basis function for which $\widehat{G}(\xi) \not> 0$ for all ξ. Define the sequence of functions $G_n(\xi) = \Lambda(\xi/n)\, G(\xi)$, $n = 1, 2, 3, \ldots$. Clearly, by Theorem 6 and the previous theorem, G_n is a basis function for $\kappa_3 < \min\{1/2, \kappa_2\}$ and $\widehat{G_n}(\xi) > 0$ for all ξ.

Further, $G_n \to G$ uniformly pointwise. This is true because the definition of Λ implies

$$|G_n(\xi) - G(\xi)| \leq \begin{cases} \dfrac{|\xi|}{n} |G(\xi)|, & |\xi| \leq n, \\ |G(\xi)|, & |\xi| > n, \end{cases}$$

and the definition of G implies $\lim\limits_{|\xi| \to \infty} G(\xi) = 0$.

A specific example is $G = \Lambda$ for which $\|G_n - G\|_\infty \leq 1/n$.

The next result gives conditions under which the convolution of two basis functions is a basis function. This is equivalent to a result showing that the product of two weight functions is a weight function.

Theorem 33 *Suppose G_1 and G_2 are the basis functions of order zero generated by the weight functions w_1 and w_2 respectively, and that each w_i has property W2 for parameter κ_i. Then the following two results hold.*

1. *If $G_1 \in L^1$, the convolution*
$$G_1 * G_2 = \frac{1}{(2\pi)^{d/2}} \int G_1 (x-y) \, G_2 (y) \, dy,$$
is the basis function generated by the weight function $w = w_1 w_2$, and w satisfies property W2 for $\kappa = \kappa_2$.

2. *If we further assume that for an integer $n_1 \leq \kappa_1$, $D^{2\alpha} G_1 \in L^1$ when $|\alpha| = n_1$, then w satisfies property W2 for parameter $\kappa = n_1 + \kappa_2$ and*
$$D^{\gamma+\delta} (G_1 * G_2) = \frac{1}{(2\pi)^{d/2}} \int D^{\gamma} G_1 (x-y) \, D^{\delta} G_2 (y) \, dy, \quad |\gamma| \leq n_1, \ |\delta| \leq \kappa_2. \tag{1.36}$$

Proof. Part 1. Define the function $w = w_1 w_2$. Clearly w is a weight function since if $\mathcal{A}_1$ and $\mathcal{A}_2$ are the weight sets, $w(x) > 0$ and $w \in C^{(0)}$ outside the set $\mathcal{A}_1 \cup \mathcal{A}_2$, which is closed and has measure zero. Since $\widehat{G_1} = \frac{1}{w_1} \in C_B^{(0)}$ we have $\frac{1}{w_1 w_2} \in L^1$ and are able to define the function $\widehat{G} = \frac{1}{w_1 w_2}$. We now show that $G = G_1 * G_2$. First observe that because $G_1 \in L^1$ and $G_2 \in C_B^{(0)}$ the convolution integral exists. Now, by definition of G_2
$$\int G_1 (x-y) \, G_2 (y) \, dy = \frac{1}{(2\pi)^{d/2}} \int \int G_1 (x-y) \frac{e^{iy\xi}}{w_2 (\xi)} d\xi \, dy,$$
and, since that this integral is absolutely convergent, we can change the order of integration so that
$$\begin{aligned} \int G_1 (x-y) \, G_2 (y) \, dy &= \frac{1}{(2\pi)^{d/2}} \int \frac{1}{w_2 (\xi)} \int e^{iy\xi} G_1 (x-y) \, dy \, d\xi \\ &= \frac{1}{(2\pi)^{d/2}} \int \frac{1}{w_2 (\xi)} \int e^{i(x-z)\xi} G_1 (z) \, dz \, d\xi \\ &= \int \frac{e^{ix\xi}}{w_2 (\xi)} \frac{1}{(2\pi)^{d/2}} \int e^{-iz\xi} G_1 (z) \, dz \, d\xi. \end{aligned}$$

The final step applies Corollary 3.7 of Petersen [11]. This states that if $f \in L^1$ and $\widehat{f} \in L^1$ then $f(x) = \frac{1}{(2\pi)^{d/2}} \int e^{ix\xi} \widehat{f}(\xi) \, d\xi$ a.e. In our case we choose $f(x) = \frac{1}{w_1(-x)}$ and obtain
$$\frac{1}{(2\pi)^{d/2}} \int G_1 (x-y) \, G_2 (y) \, dy = \frac{1}{(2\pi)^{d/2}} \int \frac{e^{ix\xi} d\xi}{w_1 (\xi) \, w_2 (\xi)} = G(x).$$

Regarding property W2, if $0 \leq \lambda \leq \kappa_2$ then, $\int \frac{|\cdot|^{2\lambda}}{w_1 w_2} \leq \left\| \frac{1}{w_1} \right\|_{\infty} \int \frac{|\cdot|^{2\lambda}}{w_2} < \infty$.

Part 2. When $|\alpha| = n_1$, we have $\widehat{D^{2\alpha} G_1} = (-1)^{|\alpha|} \xi^{2\alpha} \widehat{G_1} = (-1)^{n_1} \frac{\xi^{2\alpha}}{w_1}$ and $\widehat{D^{2\alpha} G_1} \in C_B^{(0)}$. Thus
$$\int \frac{|\cdot|^{2\kappa}}{w} = \int \frac{|\cdot|^{2n_1} |\cdot|^{2\kappa_2}}{w_1 w_2} = \sum_{|\alpha| = n_1} \frac{1}{\alpha!} \int \frac{\xi^{2\alpha}}{w_1 (\xi)} \frac{|\xi|^{2\kappa_2}}{w_2 (\xi)} d\xi < \infty,$$
and since $\frac{1}{w} \in L^1$, w satisfies W2 for $\kappa = n_1 + \kappa_2$.
Suppose that $|\gamma| \leq n_1$and $|\delta| \leq \kappa_2$. Since $G_1 \in C^{(2\kappa_1)}$, and
$\int D^{\gamma} G_1 (x-y) \, G_2 (y) \, dy$ is absolutely convergent, we have
$$D^{\gamma} (G_1 * G_2) = \frac{1}{(2\pi)^{d/2}} \int D^{\gamma} G_1 (x-y) \, G_2 (y) \, dy = \frac{1}{(2\pi)^{d/2}} \int D^{\gamma} G_1 (z) \, G_2 (x-z) \, dz.$$

Since $G_2 \in C^{(2\kappa_2)}$ and $\int D^{\gamma} G_1 (z) \, D^{\delta} G_2 (x-z) \, dz$ is absolutely convergent we have our result. ∎

1.5 The Riesz representer of the evaluation functionals $f \to (D^\alpha f)(x)$

It was noted in Remark 16 that X_w^0 is a reproducing kernel Hilbert space. However, we will use the Riesz representer of the evaluation functional $f \to f(x)$ instead of the reproducing kernel because the former can be easily extended to the evaluation of derivatives. Thus in this chapter we calculate the Riesz representers of the evaluation functionals $f \to (D^\alpha f)(x)$ and derive some of their properties. In particular, in part 6 of the next result we show that the inclusion $X_w^0 \subset C_B^{(\kappa)}$ established in Theorem 14 is continuous when $C_B^{(\kappa)}$ is endowed with the usual supremum norm. The results of this section will find use in applications where, for example, higher degrees of differentiability of a smoother are required. The main tools used here are the inverse Fourier transform formulas of Theorem 24.

Theorem 34 *Suppose the weight function w has property W2 for some κ. Then:*

1. *The unique Riesz representer $R_x \in X_w^0$ of the evaluation functional $f \to f(x)$ is*
$$R_x(z) = (2\pi)^{-d/2} G(z-x). \tag{1.37}$$
2. *If $|\alpha| \le \kappa$ then $D^\alpha R_x \in X_w^0$ and the Riesz representer for the evaluation functional $f \to (D^\alpha f)(x)$ is $(-D)^\alpha R_x$.*
3. *If $|\alpha| \le 2\kappa$ then $x \ne x'$ implies $D^\alpha R_x \ne D^\alpha R_{x'}$.*
4. *If $|\alpha| \le \kappa$ and $|\beta| \le \kappa$ then*
$$\left(D^\alpha R_x, D^\beta R_y\right)_{w,0} = (-1)^{|\beta|} \left(D^{\alpha+\beta} R_x\right)(y) = (-1)^{|\alpha+\beta|} \left(D_x^\alpha R_x, D^\beta R_y\right)_{w,0}. \tag{1.38}$$
5. *If $|\gamma| \le 2\kappa$ then*
$$\left(D^\gamma R_x\right)(y) = (-1)^{|\gamma|} \overline{\left(D^\gamma R_y\right)(x)}. \tag{1.39}$$
6. $\max\limits_{|\beta|\le\kappa} \left\|D^\beta f\right\|_\infty \le (2\pi)^{-d/4} \max\limits_{|\beta|\le\kappa} \sqrt{(-1)^{|\beta|} D^{2\beta} G(0)}\, \|f\|_{w,0}, \quad f \in X_w^0.$

Proof. Part 1. From 1.37
$$\widehat{R_x}(\xi) = (2\pi)^{-d/2} e^{-ix\xi} \widehat{G}(\xi) \tag{1.40}$$
$$= (2\pi)^{-d/2} \frac{e^{-ix\xi}}{w(\xi)}. \tag{1.41}$$

By Theorem 24, $G \in X_w^0$ and so 1.40 implies $R_x \in X_w^0$. Further, using 1.41 and then the inverse Fourier transform result 1.27 for functions in X_w^0 we obtain
$$(f, R_x)_{w,0} = \int w \widehat{f}\, \overline{\widehat{R_x}} = (2\pi)^{-d/2} \int e^{ix\xi} \widehat{f}(\xi)\, d\xi = f(x).$$

Part 2. If $|\alpha| \le \kappa$ then $\widehat{D^\alpha R_x} = (i\xi)^\alpha \widehat{R_x} = (2\pi)^{-d/2} \frac{(i\xi)^\alpha e^{-ix\xi}}{w(\xi)}$ so that
$$\|D^\alpha R_x\|_{w,0}^2 = \int w(\xi) \left|(2\pi)^{-\frac{d}{2}} \frac{(i\xi)^\alpha e^{-ix\xi}}{w(\xi)}\right|^2 d\xi \le (2\pi)^{-d} \int \frac{|\xi|^{2|\alpha|}}{w(\xi)} d\xi < \infty,$$

by weight function property W2. Further, for $f \in X_w^0$ and using the inverse Fourier transform result 1.27 for functions in X_w^0 gives
$$\begin{aligned}(f, (-D)^\alpha R_x)_{w,0} &= \int w(\xi) \widehat{f}(\xi) \overline{\widehat{(-D)^\alpha R_x}(\xi)}\, d\xi = \int w(\xi) \widehat{f}(\xi) (2\pi)^{-\frac{d}{2}} \frac{(i\xi)^\alpha e^{-ix\xi}}{w(\xi)} d\xi \\ &= (2\pi)^{-\frac{d}{2}} \int e^{-ix\xi} (i\xi)^\alpha \widehat{f}(\xi)\, d\xi \\ &= (2\pi)^{-d/2} \int e^{-ix\xi} \widehat{D^\alpha f}(\xi)\, d\xi \\ &= (D^\alpha f)(x).\end{aligned}$$

Part 3. Suppose for some $x \neq x'$ we have $R_x = R_{x'}$. Taking the Fourier transform and substituting $\widehat{G} = \frac{1}{w}$ we must have $e^{-i\xi x} = e^{-i\xi x'}$ or $e^{i\xi\left(x-x'\right)} = 1$ for almost all ξ, a contradiction.

Part 4. Since we know that $G \in C^{(2\kappa)}$ the definition of R_x implies $R_x \in C^{(2\kappa)}$ and

$$\widehat{D^\alpha R_x}\left(\xi\right) = \left(2\pi\right)^{\frac{d}{2}} \left(-i\right)^{|\alpha|} e^{-ix\xi} \xi^\alpha \widehat{G}\left(\xi\right) = \left(2\pi\right)^{\frac{d}{2}} \left(-i\right)^{|\alpha|} \frac{e^{-ix\xi}\xi^\alpha}{w\left(\xi\right)}, \quad |\alpha| \leq \kappa. \tag{1.42}$$

Next note that Theorem 24 implies that for all x, and $D^\gamma R_x \in X_w^0$ when $|\gamma| \leq \kappa$, so the first and last terms of 1.38 make sense. Now using equation 1.42

$$\begin{aligned}
\left(D^\alpha R_x, D^\beta R_y\right)_{w,0} = \int w\widehat{D^\alpha R_x}\overline{\widehat{D^\beta R_y}} &= (-1)^{|\beta|} \left(2\pi\right)^{-d} \int \left(-i\xi\right)^{\alpha+\beta} \frac{e^{i(y-x)\xi}}{w\left(\xi\right)} d\xi \\
&= (-1)^{|\beta|} \left(2\pi\right)^{-d} \int \left(-i\xi\right)^{\alpha+\beta} e^{i(y-x)\xi} \widehat{G}\left(\xi\right) d\xi \\
&= (-1)^{|\beta|} \left(2\pi\right)^{-d} \int e^{i(y-x)\xi} \widehat{D^{\alpha+\beta}G}\left(\xi\right) d\xi \\
&= (-1)^{|\beta|} \left(2\pi\right)^{-d} \left(D^{\alpha+\beta}G\right)\left(y-x\right),
\end{aligned}$$

where the last step used the inverse Fourier transform rule 1.31 for basis functions. Finally, substituting 1.37 gives $\left(D^\alpha R_x, D^\beta R_y\right)_{w,0} = (-1)^{|\beta|} \left(D^{\alpha+\beta} R_x\right)(y)$.

Regarding the second equation:

$$\begin{aligned}
\widehat{D_x^\alpha R_x}\left(\xi\right) = D_x^\alpha \widehat{G\left(\cdot - x\right)}\left(\xi\right) = (-1)^{|\alpha|} \left(D^\alpha \widehat{G)\left(\cdot - x\right)}\left(\xi\right) = (-1)^{|\alpha|} e^{-ix\xi} \widehat{D^\alpha G}\left(\xi\right) \\
= (-1)^{|\alpha|} D^\alpha \widehat{G\left(\cdot - x\right)}\left(\xi\right) \\
= (-1)^{|\alpha|} \widehat{D^\alpha R_x}\left(\xi\right),
\end{aligned}$$

so that

$$\left(D^\alpha R_x, D^\beta R_y\right)_{w,0} = (-1)^{|\alpha+\beta|} \left(D^\alpha R_x, D^\beta R_y\right)_{w,0}.$$

Part 5. If $|\gamma| \leq 2\kappa$ then $\gamma = \alpha + \beta$ for some α and β satisfying $|\alpha| \leq \kappa$ and $|\beta| \leq \kappa$. Applying part 3 twice we get

$$\begin{aligned}
D^\gamma R_x\left(y\right) = D^{\alpha+\beta} R_x\left(y\right) = (-1)^{|\beta|} \left(D^\alpha R_x, D^\beta R_y\right)_{w,0} &= (-1)^{|\beta|} \overline{\left(D^\beta R_y, D^\alpha R_x\right)}_{w,0} \\
&= (-1)^{|\alpha+\beta|} \overline{D^{\alpha+\beta} R_y\left(x\right)} \\
&= (-1)^{|\gamma|} \overline{D^\gamma R_y\left(x\right)}.
\end{aligned}$$

Part 6 From part 2, $\left\|D^\beta f\right\|_\infty = \sup\limits_{x \in \mathbb{R}^d} \left|\left(f, (-D)^\beta R_x\right)_{w,0}\right|$ when $|\beta| \leq \kappa$ so that

$$\begin{aligned}
\left\|D^\beta f\right\|_\infty \leq \sup_{x \in \mathbb{R}^d} \left\|D^\beta R_x\right\|_{w,0} \|f\|_{w,0} &= \sup_{x \in \mathbb{R}^d} \sqrt{\left(D^\beta R_x, D^\beta R_x\right)_{w,0}} \|f\|_{w,0} \\
&= \sup_{x \in \mathbb{R}^d} \sqrt{(-1)^{|\beta|} \left(D^{2\beta} R_x\right)(x)} \|f\|_{w,0}.
\end{aligned}$$

But from part 1, $\left(D^{2\beta} R_x\right)(x) = \left(2\pi\right)^{-\frac{d}{2}} D^{2\beta} G(0)$ and thus

$$\left\|D^\beta f\right\|_\infty \leq \left(2\pi\right)^{-d/4} \sqrt{(-1)^{|\beta|} D^{2\beta} G(0)} \|f\|_{w,0},$$

which implies this part. ■

1.6 More continuity properties of the data functions

The Riesz representers of the evaluation functionals $f \to D^{\alpha} f(x)$ discussed in the last section can be used to prove some local, pointwise smoothness properties of the data functions X_w^0 and their derivatives e.g. Lipschitz continuity. The basis function G generated by w also lies in X_w^0 and is considered separately. In part 2 these local properties will be used to derive the order of convergence of the variational interpolant to its data function.

1.6.1 General results

In Remark 38 we observe that the right sides of all the estimates derived in this subsection can be written in terms of even order derivatives of the basis function evaluated at the origin.

Theorem 35 *Suppose w is a weight function satisfying property W2 for parameter κ, G is the basis function and R_x is the Riesz representer of the evaluation functional $f \to f(x)$ on X_w^0. Then for $|\alpha| \leq \kappa$*

$$\|D^{\alpha}(R_x - R_y)\|_{w,0} = \frac{\sqrt{2}}{(2\pi)^{\frac{d}{4}}} \sqrt{(-1)^{|\alpha|} \left(D^{2\alpha} G(0) - \operatorname{Re}\left(D^{2\alpha} G\right)(y-x)\right)}. \tag{1.43}$$

Further, if $\kappa \geq 1$ and $|\alpha| \leq \kappa - 1$ then we have the bound

$$\|D^{\alpha}(R_x - R_y)\|_{w,0} \leq \frac{1}{(2\pi)^{\frac{d}{2}}} \left(\int \frac{\xi^{2\alpha} |\xi|^2}{w(\xi)} d\xi\right)^{\frac{1}{2}} |x-y|, \tag{1.44}$$

where

$$\int \frac{\xi^{2\alpha} |\xi|^2}{w(\xi)} d\xi = (-1)^{1+|\alpha|} (2\pi)^{\frac{d}{2}} \left(D^{2\alpha} \nabla^2 G\right)(0). \tag{1.45}$$

Proof. If $|\alpha| \leq \kappa$ then from the results of Theorem 34

$$\begin{aligned}
&\|D^{\alpha} R_x - D^{\alpha} R_y\|_{w,0}^2 \\
&= \left(D^{\alpha} R_x - D^{\alpha} R_y, D^{\alpha} R_x - D^{\alpha} R_y\right)_{w,0} \\
&= \left(D^{\alpha} R_x, D^{\alpha} R_x\right)_{w,0} - \left(D^{\alpha} R_x, D^{\alpha} R_y\right)_{w,0} - \left(D^{\alpha} R_y, D^{\alpha} R_x\right)_{w,0} + \\
&\qquad + \left(D^{\alpha} R_y, D^{\alpha} R_y\right)_{w,0} \\
&= (-1)^{|\alpha|} \left(\left(D^{2\alpha} R_x\right)(x) - \left(D^{2\alpha} R_x\right)(y) - \overline{\left(D^{2\alpha} R_x\right)(y)} + \left(D^{2\alpha} R_y\right)(y)\right) \\
&= (-1)^{|\alpha|} (2\pi)^{-\frac{d}{2}} \left(D^{2\alpha} G(0) - \left(D^{2\alpha} G\right)(y-x) - \overline{\left(D^{2\alpha} G\right)(y-x)} + D^{2\alpha} G(0)\right) \\
&= (-1)^{|\alpha|} (2\pi)^{-\frac{d}{2}} \left(D^{2\alpha} G(0) - \left(D^{2\alpha} G\right)(y-x) - \overline{\left(D^{2\alpha} G\right)(y-x)} + D^{2\alpha} G(0)\right) \\
&= (-1)^{|\alpha|} 2 (2\pi)^{-\frac{d}{2}} \left(D^{2\alpha} G(0) - \operatorname{Re}\left(D^{2\alpha} G\right)(y-x)\right).
\end{aligned}$$

The proof of our second result uses equation 1.42 i.e. $\widehat{D^{\alpha} R_x}(\xi) = \frac{(-i)^{|\alpha|}}{(2\pi)^{\frac{d}{2}}} \frac{e^{-ix\xi} \xi^{\alpha}}{w(\xi)}$, so that

$$\begin{aligned}
\|D^{\alpha} R_x - D^{\alpha} R_y\|_{w,0}^2 = \int w \left|\widehat{D^{\alpha} R_x - D^{\alpha} R_y}\right|^2 &= \frac{1}{(2\pi)^d} \int \left|e^{i\xi x} - e^{i\xi y}\right|^2 \frac{\xi^{2\alpha}}{w(\xi)} d\xi \\
&= \frac{1}{(2\pi)^d} \int \left(2 \sin\left(\frac{(x-y)\xi}{2}\right)\right)^2 \frac{\xi^{2\alpha}}{w(\xi)} d\xi \\
&= \frac{1}{(2\pi)^d} \int |(x-y)\xi|^2 \left(\frac{\sin((x-y)\xi/2)}{(x-y)\xi/2}\right)^2 \frac{\xi^{2\alpha}}{w(\xi)} d\xi \\
&\leq \frac{1}{(2\pi)^d} \left(\int \frac{\xi^{2\alpha} |\xi|^2}{w(\xi)} d\xi\right) |x-y|^2,
\end{aligned}$$

which is finite since $1 + |\alpha| \leq \kappa$. Continuing

$$\int \frac{\xi^{2\alpha} |\xi|^2}{w(\xi)} d\xi = \int \xi^{2\alpha} |\xi|^2 \widehat{G}(\xi) d\xi = (-1)^{1+|\alpha|} \int \widehat{D^{2\alpha} \nabla^2 G} = (2\pi)^{\frac{d}{2}} (-1)^{1+|\alpha|} \left(D^{2\alpha} \nabla^2 G\right)(0).$$

■
We now prove some uniform pointwise estimates for functions in X_w^0.

Corollary 36 *Suppose the weight function w has property W2 for parameter κ and that G is the basis function of order 0. Then we have the following local pointwise estimates for functions $f \in X_w^0$:*

1. *If $\kappa \geq 1$ and $|\alpha| \leq \kappa - 1$ then*

$$|D^\alpha f(x) - D^\alpha f(y)| \leq \frac{1}{(2\pi)^{\frac{d}{2}}} \|f\|_{w,0} \left(\int \frac{|\xi|^2 \xi^{2\alpha}}{w(\xi)} d\xi \right)^{\frac{1}{2}} |x - y| .$$

2. *If $|\alpha| \leq \kappa$ then*

$$|D^\alpha f(x) - D^\alpha f(y)| \leq \frac{\sqrt{2}}{(2\pi)^{\frac{d}{4}}} \|f\|_{w,0} \sqrt{(-1)^{|\alpha|} \left(D^{2\alpha} G(0) - \mathrm{Re}\left(D^{2\alpha} G \right)(y - x) \right)}. \tag{1.46}$$

3. *For the hat weight function in one dimension $\kappa < 1/2$ and*

$$|f(x) - f(y)| \leq \frac{\sqrt{2}}{(2\pi)^{\frac{d}{4}}} \|f\|_{w,0} |x - y|^{\frac{1}{2}}, \quad x, y \in \mathbb{R}^1, \ |x - y| \leq 1.$$

Proof. Part 1 This is an application of inequality 1.44 of Theorem 35

$$\begin{aligned} |D^\alpha f(x) - D^\alpha f(y)| &= \left| (f, D^\alpha R_x - D^\alpha R_y)_{w,0} \right| \leq \|f\|_{w,0} \|D^\alpha R_x - D^\alpha R_y\|_{w,0} \\ &\leq \frac{1}{(2\pi)^{\frac{d}{2}}} \|f\|_{w,0} \left(\int \frac{|\xi|^2 \xi^{2\alpha}}{w(\xi)} d\xi \right)^{\frac{1}{2}} |x - y| . \end{aligned}$$

Part 2 By part 2 of Theorem 34

$$|D^\alpha f(x) - D^\alpha f(y)| = \left| (f, D^\alpha R_x - D^\alpha R_y)_{w,0} \right| \leq \|f\|_{w,0} \|D^\alpha R_x - D^\alpha R_y\|_{w,0} ,$$

and inequality 1.43 of Corollary 35 completes the proof.

Part 3 From Theorem 6 $\kappa < 1/2$ so $\alpha = 0$ and since $G(x) = 1 - |x|$ when $|x| \leq 1$ our result follows from the bound proved in part 2. ■

The next corollary provides information concerning the smoothness properties of basis functions.

Corollary 37 *Suppose G is the basis function generated by a weight function w satisfying property W2 for parameter κ. Then for each $z \in \mathbb{R}^d$, G has the following properties:*

1. *If $|\alpha| \leq \kappa$ and $|\beta| \leq \kappa$ then*

$$\left| D^{\alpha+\beta} G(x) - D^{\alpha+\beta} G(y) \right| \leq k \sqrt{(-1)^{|\alpha|} \left(D^{2\alpha} G(0) - \mathrm{Re}\left(D^{2\alpha} G \right)(x - y) \right)}, \tag{1.47}$$

where

$$k = \frac{\sqrt{2}}{(2\pi)^{\frac{d}{4}}} \left(\int \frac{\xi^{2\alpha} d\xi}{w(\xi)} \right)^{\frac{1}{2}} = \sqrt{2} \sqrt{(-1)^{|\alpha|} D^{2\alpha} G(0)}. \tag{1.48}$$

2. *If $|\alpha| \leq \kappa$ then*

$$\left| D^{2\alpha} G(x) \right| \leq (-1)^{|\alpha|} D^{2\alpha} G(0) ,$$

and

$$(-1)^{|\alpha|} \mathrm{Re}\, D^{2\alpha} G(x) < (-1)^{|\alpha|} D^{2\alpha} G(0) , \quad x \neq 0.$$

3. *Suppose* $\kappa \geq 1$. *Then if* $|\alpha| \leq \kappa - 1$ *and* $|\beta| \leq \kappa - 1$

$$\left|D^{\alpha+\beta}G(x) - D^{\alpha+\beta}G(y)\right| \leq \tfrac{1}{(2\pi)^{\frac{d}{2}}} \left(\int \frac{\xi^{2\beta} d\xi}{w(\xi)}\right)^{\frac{1}{2}} \left(\int \frac{|\xi|^2 \xi^{2\alpha} d\xi}{w(\xi)}\right)^{\frac{1}{2}} |x-y|.$$

Proof. Part 1. Our starting point is equation 1.38 i.e. for $x, y \in \mathbb{R}^d$, $\left(D^\alpha R_x, D^\beta R_y\right)_{w,0} = (-1)^{|\beta|}\left(D^{\alpha+\beta} R_x\right)(y)$ when $|\alpha| \leq \kappa$, $|\beta| \leq \kappa$. Hence, since $R_x(y) = (2\pi)^{-\frac{d}{2}} G(y-x)$

$$\begin{aligned} D_z^{\alpha+\beta}\left(G(z-x) - G(z-y)\right) &= (2\pi)^{-\frac{d}{2}} D^{\alpha+\beta}\left(R_x - R_y\right)(z) \\ &= (2\pi)^{-\frac{d}{2}} \left(D^\alpha\left(R_x - R_y\right), D^\beta R_z\right)_{w,0}, \end{aligned}$$

so that

$$\left|D_z^{\alpha+\beta}\left(G(z-x) - G(z-y)\right)\right| \leq (2\pi)^{-\frac{d}{2}} \left\|D^\alpha\left(R_x - R_y\right)\right\|_{w,0} \left\|D^\beta R_z\right\|_{w,0}. \qquad (1.49)$$

From equation 1.42 we conclude that $\left\|D^\beta R_z\right\|_{w,0} = (2\pi)^{-\frac{d}{2}} \left(\int \frac{\xi^{2\beta} d\xi}{w(\xi)}\right)^{\frac{1}{2}}$ and then using equation 1.43 for $\left\|D^\alpha\left(R_x - R_y\right)\right\|_{w,0}$ inequality 1.47 follows.

Part 2. We have $(-1)^{|\alpha|} D^{2\alpha}G(x) = (2\pi)^{-\frac{d}{2}} \int e^{i\xi x} \frac{\xi^{2\alpha}}{w(\xi)} d\xi$, so that

$$\left|D^{2\alpha}G(x)\right| \leq (2\pi)^{-\frac{d}{2}} \int \frac{\xi^{2\alpha}}{w(\xi)} d\xi = (-1)^{|\alpha|} D^{2\alpha}G(0).$$

From part 1 we know that $(-1)^{|\alpha|} \operatorname{Re}\left(D^{2\alpha}G\right)(x-y) \leq (-1)^{|\alpha|} D^{2\alpha}G(0)$. Now suppose that $(-1)^{|\alpha|} \operatorname{Re}\left(D^{2\alpha}G\right)(z) = (-1)^{|\alpha|} D^{2\alpha}G(0)$ for some $z \neq 0$. Equation 1.43 would then imply that $D^\alpha R_x = D^\alpha R_{z+x}$ for all x, and hence $z = 0$ by part 3 of Theorem 34.

Part 3. Our starting point is inequality 1.49 of part 1. Substitute the expression for $\left\|D^\beta R_z\right\|_{w,0}$ given in the proof of part 1 and then use inequality 1.44 to estimate $\left\|D^\alpha\left(R_x - R_y\right)\right\|_{w,0}$. ■

Remark 38 *Observe that by using 1.45 and 1.48 the right sides of all the estimates involving weight functions in this subsection can be written in terms of even order derivatives of the basis function evaluated at the origin.*

1.6.2 Better results obtained using Taylor series

We can improve on part 3 of Corollary 37 for the special case $\kappa \geq 1$, $y = 0$ and $\alpha = \beta = 0$; improved in the sense that $|x|$ is replaced by $|x|^2$ and this is done by using the Taylor series expansion with integral remainder given in Subsection A.7 of the Appendix:

Theorem 39 *Suppose the weight function* w *satisfies property W2 for* $\kappa = 1$. *Then* $\left(\nabla^2 G\right)(0)$ *is real and negative and we have the bound*

$$G(0) - \operatorname{Re} G(x) \leq -\frac{d}{2}\left(\nabla^2 G\right)(0)|x|^2, \quad x \in \mathbb{R}^d. \qquad (1.50)$$

Now suppose w *is radial and let* $r = |x|$. *Then:*

1. *If* $G(x) = g(r)$ *then* $g \in C_B^{(2)}([0,\infty))$ *and* $r^{-1}g' \in C_B^{(0)}([0,\infty))$. *Also*

$$\left(\nabla^2 G\right)(0) = g''(0)\, d. \qquad (1.51)$$

2. *If* $G(x) = f\left(r^2\right)$ *then* $f \in C^{(2)}((0,\infty)) \cap C_B^{(1)}([0,\infty))$ *and* $rf''(r) \in C_B^{(0)}([0,\infty))$ *and* $\lim\limits_{r \to 0^+} rf''(r) = 0$. *Also*

$$\left(\nabla^2 G\right)(0) = 2f'(0)\, d. \qquad (1.52)$$

Proof. We start by assuming w is an even function. Because $\kappa = 1$ it follows that $G \in C_B^{(\lfloor 2\kappa \rfloor)} \subset C_B^{(2)}$.
Also, since

$$D^\beta G(x) = (2\pi)^{-\frac{d}{2}} \int e^{-ix\xi} \frac{(i\xi)^\beta}{w(\xi)} d\xi, \quad |\beta| \leq 2,$$

and w is even, it follows that G is real, $|G(x)| \leq G(0)$ and $D_k G(0) = 0$ for all k. Thus from Appendix A.7 we now have the second order Taylor series expansion

$$G(x) = G(0) + (\mathcal{R}_2 G)(0, x),$$

with remainder estimate

$$|(\mathcal{R}_2 G)(0, x)| \leq \frac{d}{2} \left(\max_{\substack{|\beta|=2 \\ t \in [0,x]}} \left| D^\beta G(t) \right| \right) |x|^2, \quad x \in \mathbb{R}^d.$$

But when $|\beta| = 2$, $D^\beta G(x) = -(2\pi)^{-\frac{d}{2}} \int e^{-ix\xi} \frac{\xi^\beta}{w(\xi)} d\xi$ and hence

$$\left| D^\beta G(x) \right| \leq (2\pi)^{-\frac{d}{2}} \int \frac{|\xi|^{|\beta|}}{w(\xi)} d\xi = (2\pi)^{-\frac{d}{2}} \int \frac{|\xi|^2}{w(\xi)} d\xi = -\left(\nabla^2 G\right)(0),$$

so that

$$G(0) - G(x) \leq -\frac{d}{2} \left(\nabla^2 G\right)(0) |x|^2, \quad x \in \mathbb{R}^d. \tag{1.53}$$

Now set aside the assumption that w is even. We define the even function w_e by

$$\frac{1}{w_e(\xi)} = \frac{1}{2} \left(\frac{1}{w(\xi)} + \frac{1}{w(-\xi)} \right), \tag{1.54}$$

and it is easy to show that w_e has property W1 and satisfies

$$\int \frac{|\xi|^\lambda}{w_e(\xi)} d\xi = \int \frac{|\xi|^\lambda}{w(\xi)} d\xi, \quad 0 \leq \lambda \leq \kappa, \tag{1.55}$$

so that w_e has property W2 for κ. Further, if w_e has basis function G_e then $\widehat{G_e} = \frac{1}{2}\left(\widehat{G} + \widehat{\overline{G}}\right)$ and hence $G_e = \operatorname{Re} G$. But G_e satisfies 1.53 and since 1.55 implies $\left(\nabla^2 G_e\right)(0) = \left(\nabla^2 G\right)(0)$ we can conclude that 1.50 is true.

$\boxed{\text{Part 1}}$ Suppose $G(x) = g(r)$. Then substituting $x = (r, 0')$ we have

$$g(r) = G(r, 0'),$$

so that it is clear that $g \in C_B^{(2)}([0,\infty))$ since $G \in C_B^{(2)}$. Regarding $r^{-1}g'$, it is clear that $r^{-1}g' \in C_B^{(2)}((0,\infty))$. Further, since $g' \in C_B^{(1)}([0,\infty))$, $\lim\limits_{r \to 0^+} \frac{g'(r)}{r} = g''(0)$ and $\lim\limits_{r \to \infty} \frac{g'(r)}{r} = 0$ which means that $r^{-1}g' \in C_B^{(2)}([0,\infty))$.
If $x \neq 0$

$$D_k G(x) = \frac{x_k}{r} g'(r),$$

and

$$D_j D_k G(x) = \frac{x_j x_k}{r^2} g''(r) - \frac{x_j x_k}{r^3} g'(r) + \frac{\delta_{j,k}}{r} g'(r),$$

so that

$$\nabla^2 G(x) = \sum_{k=1}^d D_k^2 G(x) = g''(r) - \frac{g'(r)}{r} + d\frac{g'(r)}{r} = g''(r) + (d-1)\frac{g'(r)}{r}.$$

But $g \in C_B^{(2)}([0,\infty))$ so

$$\lim_{x \to 0} \nabla^2 G(x) = g''(0) + (d-1) g''(0) = g''(0) d.$$

$\boxed{\text{Part 2}}$ Suppose $G(x) = f(r^2)$. Then the equation $f(r) = g(\sqrt{r})$ and the properties of f proved in part 1 can be used to easily derive the results of this part. ■

For radial basis functions the estimates the last theorem for $G(0) - \operatorname{Re} G(x)$ give a d^2 dependency. However, if the basis function is a radial function this dependency can be avoided:

Theorem 40 *Suppose the weight function w satisfies property W2 for $\kappa = 1$ and denote the basis function by G. Then:*

1. *If $G(x) = f(r^2)$ then*

$$G(0) - \operatorname{Re} G(x) \le \|2rf'' + f'\|_\infty |x|^2, \quad x \in \mathbb{R}^d. \tag{1.56}$$

2. *If $G(x) = g(r)$ then*

$$G(0) - \operatorname{Re} G(x) \le \frac{1}{2} \|g''\|_\infty |x|^2, \quad x \in \mathbb{R}^d, \tag{1.57}$$

where $\|2rf'' + f'\|_\infty = \frac{1}{2}\|g''\|_\infty$.

Proof. We will start by assuming w is an even function. Because $\kappa \ge 1$ it follows that $G \in C_B^{(\lfloor 2\kappa \rfloor)} \subset C_B^{(2)}$. Also, since

$$D^\beta G(x) = (2\pi)^{-\frac{d}{2}} \int e^{-ix\xi} \frac{(i\xi)^\beta}{w(\xi)} d\xi, \quad |\beta| \le 2,$$

it follows that if w is even then G is real, $|G(x)| \le G(0)$ and $D_k G(0) = 0$ for all k. Thus from Appendix A.7 we now have the Taylor series expansion

$$G(x) = G(0) + (\mathcal{R}_2 G)(0, x),$$

where

$$(\mathcal{R}_2 G)(0, x) = 2 \sum_{|\beta|=2} \frac{x^\beta}{\beta!} \int_0^1 (1-t) \left(D^\beta G\right)(tx)\, dt. \tag{1.58}$$

$\boxed{\text{Part 1}}$ Suppose $G(x) = f(r^2)$. Then

$$D_k G(x) = 2x_k f'(r^2),$$

and

$$D_j D_k G(x) = 4x_j x_k f''(r^2) + 2\delta_{j,k} f'(r^2).$$

Set $\delta_\beta = \delta_{2,\max\beta}$. Then in multi-index notation

$$D^\beta G(x) = 4x^\beta f''(r^2) + 2\delta_\beta f'(r^2), \quad |\beta| = 2,$$

and so, using the identity A.1, the remainder can be written

$$\begin{aligned}
(\mathcal{R}_2 G)(0,x) &= 2\sum_{|\beta|=2}\frac{x^\beta}{\beta!}\int_0^1 (1-t)\left(4x^\beta f''\left(r^2\right)\right)(tx)\,dt+ \\
&\quad + 2\sum_{|\beta|=2}\frac{x^\beta}{\beta!}\int_0^1 (1-t)\left(2\delta_\beta f'\left(r^2\right)\right)(tx)\,dt \\
&= 8\sum_{|\beta|=2}\frac{x^\beta}{\beta!}\int_0^1 (1-t)\left(t^2x^\beta\right) f''\left(t^2r^2\right)dt+ \\
&\quad + 4\sum_{|\beta|=2}\frac{x^\beta}{\beta!}\int_0^1 (1-t)\,\delta_\beta f'\left(t^2r^2\right)dt \\
&= 8\left(\sum_{|\beta|=2}\frac{x^{2\beta}}{\beta!}\right)\int_0^1 (1-t)\,t^2 f''\left(t^2r^2\right)dt+ \\
&\quad + 4\left(\sum_{|\beta|=2}\delta_\beta\frac{x^\beta}{\beta!}\right)\int_0^1 (1-t)\,f'\left(t^2r^2\right)dt \\
&= 4r^4\int_0^1 (1-t)\,t^2 f''\left(t^2r^2\right)dt + 2r^2\int_0^1 (1-t)\,f'\left(t^2r^2\right)dt \\
&= r^2\int_0^1 (1-t)\left(4t^2r^2 f''\left(t^2r^2\right)+2f'\left(t^2r^2\right)\right)dt.
\end{aligned} \tag{1.59}$$

The remainder can be estimated as

$$\begin{aligned}
\left|(\mathcal{R}_2 G)(0,x)\right| &\le r^2\int_0^1 (1-t)\left|4t^2r^2 f''\left(t^2r^2\right)+2f'\left(t^2r^2\right)\right|dt \\
&= r^2\left\|4rf''+2f'\right\|_\infty\int_0^1 (1-t)\,dt \\
&= \left\|2rf''+f'\right\|_\infty |x|^2 .
\end{aligned} \tag{1.60}$$

so that we have proved the desired estimate 1.56 when w is an even function. This estimate can now be extended to an arbitrary weight function by the technique used in Theorem 39 which involved defining the even weight function 1.54.

$\boxed{\text{Part 2}}$ Suppose $G(x) = g(r)$. Then $g(r) = f\left(r^2\right)$ and

$$g''(r) = 2f'\left(r^2\right) + 4r^2 f''\left(r^2\right), \tag{1.61}$$

so the remainder equation 1.59 can be written

$$(\mathcal{R}_2 G)(0,x) = r^2\int_0^1 (1-t)\left(4t^2r^2 f''\left(t^2r^2\right)+2f'\left(t^2r^2\right)\right)dt = \frac{1}{2}r^2\int_0^1 (1-t)\,g''(tr)\,dt,$$

and estimated by

$$\left|(\mathcal{R}_2 G)(0,x)\right| \le \frac{1}{2}\left\|g''\right\|_\infty |x|^2 . \tag{1.62}$$

Equation 1.61 ensures that the estimates 1.60 and 1.62 are equal.

We have now proved the desired estimate 1.57 when w is an even function. This estimate can now be extended to an arbitrary weight function by the technique of Theorem 39 which involved defining the even weight function 1.54. ■

Remark 41 *Calculations using part 1 may be easier when the basis function depends on r^2: see the Type 1 convergence estimates for the radial functions of Subsubsection 2.5.2.*

Inequality 1.47 shows that the smoothness of the basis function near the origin implies its global smoothness. Specifically, for the one-dimensional hat function Λ we have $\kappa < 1/2$ and near the origin $|\Lambda(x) - \Lambda(y)| \leq \sqrt{2}\,|x-y|^{1/2}$ by part 3 of Corollary 36. However, Λ actually satisfies the stronger estimate $|\Lambda(x) - \Lambda(y)| \leq |x-y|$ everywhere i.e. it is uniformly Lipschitz continuous on $\mathbb{R}^1$ and Theorem 43 will generalize this result. First we will need the following extension to distributions of the Taylor series expansion with integral remainder given in Appendix A.7.

Lemma 42 *Suppose $u \in C_B^{(0)}\left(\mathbb{R}^d\right)$ and the distributional derivatives $\left\{D^\beta u\right\}_{|\beta|=1}$ are bounded functions. Then*

$$u(z+b) = u(z) + \left(\mathcal{R}_1 u\right)(z,b),$$

where $\mathcal{R}_1 u$ is the integral remainder term

$$\left(\mathcal{R}_1 u\right)(z,b) = \sum_{|\beta|=1} \frac{b^\beta}{\beta!} \int_0^1 (D^\beta u)(z + (1-t)\,b)dt,$$

which satisfies

$$\left|\left(\mathcal{R}_1 u\right)(z,b)\right| \leq \sqrt{d} \max_{|\beta|=1} \left\|D^\beta u\right\|_\infty |b|. \tag{1.63}$$

Proof. In order to overcome the fact that $D^\beta u$ may not be $C^{(0)}\left(\mathbb{R}^d\right)$ when $|\beta| = 1$, we will use a Taylor series expansion with remainder for distributions. Suppose $\phi \in C_0^\infty$. Then the conditions on u render all the integrals absolutely convergent and allows us to apply Fubini's theorem to swap the order of integration in the calculations of this lemma:

$$\begin{aligned}
[u(z+b), \phi(z)] &= [u(z), \phi(z-b)] \\
&= [u(z), \phi(z)] + \left[u(z), \sum_{|\beta|=1} \frac{(-b)^\beta}{\beta!} \int_0^1 t^{2\kappa} \left(D^\beta \phi\right)(z + (1-t)(-b))\,dt\right] \\
&= [u(z), \phi(z)] - \sum_{|\beta|=1} \frac{b^\beta}{\beta!} \left[u(z), \int_0^1 \left(D^\beta \phi\right)(z - (1-t)\,b)\,dt\right].
\end{aligned} \tag{1.64}$$

We now analyze the integral remainder term of 1.64. When $|\beta| = 1$

$$\begin{aligned}
\left[u(z), \int_0^1 \left(D^\beta \phi\right)(z-(1-t)\,b)\,dt\right] &= \int u(z) \int_0^1 t^n \left(D^\beta \phi\right)(z-(1-t)\,b)\,dt\ dz \\
&= \int_0^1 \int u(z) \left(D^\beta \phi\right)(z-(1-t)\,b)\,dz\ dt \\
&= \int_0^1 \left[u(z), \left(D^\beta \phi\right)(z-(1-t)\,b)\right]\ dt \\
&= \int_0^1 \left[u(z+(1-t)\,b), D^\beta \phi(z)\right]\ dt \\
&= (-1)^{|\beta|} \int_0^1 \left[\left(D^\beta u\right)(z+(1-t)\,b), \phi(z)\right]\ dt \\
&= -\int_0^1 \int \left(D^\beta u\right)(z+(1-t)\,b)\,\phi(z)\,dz\ dt \\
&= -\int \int_0^1 \left(D^\beta u\right)(z+(1-t)\,b)\,dt\ \phi(z)\,dz \\
&= -\left[\int_0^1 \left(D^\beta u\right)(z+(1-t)\,b)\,dt, \phi(z)\right],
\end{aligned}$$

so 1.64 now becomes

$$[u(z+b), \phi(z)] = [u(z), \phi(z)] - \sum_{|\beta|=1} \frac{b^\beta}{\beta!} \left[\int_0^1 \left(D^\beta u\right)(z+(1-t)\,b)\,dt, \phi(z)\right],$$

for $\phi \in C_0^\infty$. Thus

$$u(z+b) = u(z) + \sum_{|\beta|=1} \frac{b^\beta}{\beta!} \int_0^1 \left(D^\beta u\right)(z+(1-t)b)\,dt,$$

as claimed. Finally, the estimate 1.63 follows directly from the integral remainder estimate A.3 of the Appendix. ■

We now prove that all derivatives of the extended B-spline tensor product basis function up to order $2n-2$ are uniformly Lipschitz continuous of order 1.

Theorem 43 *Let $G(x) = \prod_{k=1}^{d} G_1(x_k)$ be an extended B-spline tensor product basis function, as introduced in Subsection 1.4.4. We then have the estimates*

$$|G(x) - G(y)| \le \sqrt{d} G_1(0)^{d-1} \|DG_1\|_\infty |x-y|, \quad x, y \in \mathbb{R}^d,$$

and if $|\alpha| \le 2n-2$

$$|D^\alpha G(x) - D^\alpha G(y)| \le \sqrt{d} \max_{k=1}^{d} \|D_k D^\alpha G\|_\infty |x-y|. \tag{1.65}$$

Proof. From Theorem 29 we know that $G \in C_0^{(2n-2)}\left(\mathbb{R}^d\right)$ and that the derivatives $\left\{D^\beta G\right\}_{|\beta|=2n-1}$ are bounded functions. Consequently $G \in C_0^{(0)}\left(\mathbb{R}^d\right)$ and all the first derivatives are bounded functions. This implies $u = G$ satisfies the conditions of Lemma 42 with $z = y$ and $b = x - y$ so the estimate of that lemma holds i.e.

$$|G(x) - G(y)| \le \sqrt{d} \max_{|\beta|=1} \left\|D^\beta G\right\|_\infty |x-y|. \tag{1.66}$$

Again by Theorem 29, $G_1 \in C_0^{(0)}\left(\mathbb{R}^1\right)$ and DG_1 is a bounded function so that

$$|G(x) - G(y)| \le \sqrt{d} \|G_1\|_\infty^{d-1} \|DG_1\|_\infty |x-y|.$$

Finally, part 2 of Corollary 37 tells us that $\|G_1\|_\infty \le G_1(0)$ which completes the proof of the first inequality. To prove the second inequality we simply replace G by $D^\alpha G$ in 1.66. ■

2

The minimal norm interpolant

2.1 Introduction

In Chapter 1 of this document we introduced the basic mathematical machinery: weight function, data space and basis function. In this chapter this theory is applied to the well-known minimal norm interpolation problem. This problem is solved and several pointwise convergence results are derived. These are illustrated with numerical results obtained using several extended B-spline basis functions and special classes of data functions.

In more detail, the functions from the Hilbert data space X_w^0 are used to define the standard minimal norm interpolation problem with independent data $X = \left\{x^{(i)}\right\}_{i=1}^N$ and dependent data $y = \{y_i\}_{i=1}^N$, the latter being obtained by evaluating a data function at X. This problem is then shown to have a unique basis function solution of the form $\sum\limits_{i=1}^N \alpha_i G\left(\cdot - x^{(i)}\right)$ with $\alpha_i \in \mathbb{C}$. The standard matrix equation for the α_i is then derived.

We will then consider several categories of estimates for the pointwise convergence of the interpolant to its data function when the data is confined to a bounded data region. In all these results the order of convergence appears as the power of the radius of the largest ball in the data region that can be fitted between the X data points. The convergence results of this document can be divided into those which use Lagrange interpolation theory, and thus assume unisolvent data X (Definition 66), and those which do not. The non-unisolvency proofs are *much* simpler than the unisolvency-based results and if the data functions are chosen appropriately the constants can be calculated. However, the maximum order of convergence obtained is 1, no matter how large the parameter κ. On the other hand, the constants for the unisolvency-based results are more difficult to calculate but the order of convergence is equal to the value of max $\lfloor \kappa \rfloor$.

Non-unisolvent data: Type 1 interpolation error estimates Suppose that the weight function has property W2 for some $\kappa \geq 0$, that the data X is contained in a closed, bounded, infinite data set K, and that the basis function G satisfies the estimate

$$|G(0) - \operatorname{Re} G(x)| \leq C_G |x|^{2s}, \quad |x| < h_G,$$

for some $h_G > 0$. Then in Theorem 59 it is shown that the interpolant $\mathcal{I}_X f$ of a data function $f \in X_w^0$ satisfies

$$|f(x) - \mathcal{I}_X f(x)| \leq k_G \|f\|_{w,0} (h_{X,K})^s, \quad x \in K,$$

when $h_{X,K} = \max\limits_{x \in K} \operatorname{dist}(x, X) \leq h_G$ and $k_G = (2\pi)^{-\frac{d}{4}} \sqrt{2C_G}$. This implies an order of convergence of at least s.

$\boxed{\text{Non-unisolvent data: Type 2 interpolation error estimates}}$ Here we avoid making any assumptions about G and instead assume $\kappa \geq 1$ and use the properties of the Riesz representer R_x of the evaluation functional $f \to f(x)$. A consequence of this approach is the estimate (Theorem 65): when $h_{X,K} < \infty$

$$|f(x) - \mathcal{I}_X f(x)| \leq k_G \|f\|_{w,0} h_{X,K}, \quad x \in K, \ f \in X_w^0,$$

where

$$k_G = (2\pi)^{-d/4} \sqrt{-\left(\nabla^2 G\right)(0)}.$$

so the order of convergence is **always** at least 1.

$\boxed{\text{Unisolvent data - interpolation error estimates}}$ Here X is assumed to have a unisolvent subset of order max $\lfloor\kappa\rfloor$ and is contained in an open, bounded data region Ω. In this case it follows from Theorem 74 that there exist constants $h_{\Omega,\kappa}, k_G > 0$ such that

$$|f(x) - \mathcal{I}_X f(x)| \leq k_G \|f\|_{w,0} (h_{X,\Omega})^{\lfloor\kappa\rfloor}, \quad x \in \overline{\Omega}, \ f \in X_w^0,$$

when $h_{X,\Omega} = \sup\limits_{\omega\in\Omega} \operatorname{dist}(\omega, X) < h_{\Omega,\kappa}$.

Numerical results are only presented for the non-unisolvent data cases and these illustrate the convergence of the interpolant to it's data function. We will only be interested in the convergence of the interpolant to it's data function and not in the algorithm's performance as an interpolant. Only the extended B-splines will be considered and we will also restrict ourselves to one dimension so that the data density parameter $h_{X,\Omega}$ can be easily calculated. If we can calculate the data function norm we can calculate the error estimate. All the extended B-spline basis weight functions have a power of $\sin^2 x$ in the denominator and so we have derived special classes of data functions for which the data function norm can be calculated. The derivation of these special classes of data functions led to the characterization of the restriction spaces $X_w^0(\Omega)$ for various class of weight function. For example, Theorem 90 shows that the restrictions of the B-spline data functions are a member of the class of *Sobolev-like spaces*:

$$H_s^m(\Omega) = \left\{u \in L^2(\Omega) : D^\alpha u \in L^2(\Omega) \ \ for \ \alpha_i \leq m, \ 0 \leq i \leq m\right\}, \quad m = 1, 2, 3, \ldots$$

It is easy to construct functions that are in these spaces.

As expected interpolant instability is evident and because our error estimates assume an infinite precision we filter the error to remove spikes which are a manifestation of the instability.

2.2 The space $W_{G,X}$

In this section we will define the space $W_{G,X}$ which contains the solution to the minimal norm interpolation problem 2.4. Here X is the independent data and G is the basis function. The next two results will be required to show that $W_{G,X}$ is a well defined finite dimensional vector space. The next theorem requires the following lemma which we state without proof.

Lemma 44 *Suppose* $\left\{x^{(k)}\right\}_{k=1}^N$ *is a set of distinct points in* $\mathbb{R}^d$ *and* $v = (\nu_k)_{k=1}^N \in \mathbb{C}^N$*. Define the function* a_v *by*

$$a_v(\xi) = \sum_{k=1}^N \nu_k e^{-ix^{(k)}\xi}.$$

Then:

1. *if* $a_v(\xi) = 0$ *for all* ξ *then* $v_k = 0$ *for all* k.
2. *The null space of* a_v *is a closed set of measure zero.*

Theorem 45 *Let* $X = \left\{x^{(k)}\right\}_{k=1}^N$ *be a set of distinct points in* $\mathbb{R}^d$*. Then the set of translated basis functions* $\left\{G\left(\cdot - x^{(k)}\right)\right\}_{k=1}^N$ *is linearly independent with respect to the complex scalars.*

Proof. Suppose for complex α_k, $\sum_{k=1}^{N} \alpha_k G\left(x - x^{(k)}\right) = 0$. Taking the Fourier transform we obtain

$$\frac{1}{w(\xi)} \sum_{k=1}^{N} \alpha_k e^{ix^{(k)}\xi} = 0 \quad a.e.$$

Since $w(\xi) > 0$ a.e. we have $\sum_{k=1}^{N} \alpha_k e^{ix^{(k)}\xi} = 0$ and part 1 of Lemma 44 implies $\alpha_k = 0$ for all k. ■

We now introduce the space $W_{G,X}$.

Definition 46 ***The finite dimensional vector space*** $W_{G,X}$

Suppose the weight function w has property W2 for parameter κ, let G be the (continuous) basis function generated by w and let $X = \left\{x^{(k)}\right\}_{k=1}^{N}$ be a set of distinct points in $\mathbb{R}^d$. We have shown in Theorem 45 that the functions $G\left(\cdot - x^{(k)}\right)$ are linearly independent so the span of these functions makes sense. We define the $N-$ dimensional vector space $W_{G,X}$ by

$$W_{G,X} = \left\{ \sum_{k=1}^{N} \alpha_k G\left(\cdot - x^{(k)}\right) : \alpha_k \in \mathbb{C} \right\}. \tag{2.1}$$

When convenient below, functions in $W_{G,X}$ will be written in the form $f_\alpha(x) = \sum_{k=1}^{N} \alpha_k G\left(x - x^{(k)}\right)$ where $\alpha = (\alpha_k) \in \mathbb{C}^N$.

2.3 The matrices $G_{X,X}$ and $R_{X,X}$

We will now define the *basis function matrix* $G_{X,X}$ and the *reproducing kernel matrix* $R_{X,X}$. In this document we deal with basis functions of order zero so we will deduce the simple relationship $R_{X,X} = (2\pi)^{-d/2} G_{X,X}$. These matrices will be used to construct the matrix equations for the interpolants studied in this document and the smoothers studied in Chapters 4 and 5.

Definition 47 ***The basis function matrix*** $G_{X,X}$

Let $X = \left\{x^{(n)}\right\}_{n=1}^{N}$ be a set of distinct points in $\mathbb{R}^d$ and suppose G is the basis function generated by w. Then the basis function matrix $G_{X,X}$ is defined by

$$G_{X,X} = \left(G\left(x^{(i)} - x^{(j)}\right)\right).$$

Remark 48 *Since $G = \left(\frac{1}{w}\right)^{\vee}$ we have $\overline{G(-x)} = G(x)$ and so the matrix $G_{X,X}$ is Hermitian.*

Definition 49 ***The reproducing kernel matrix*** $R_{X,X}$

Suppose R_x is the Riesz representer of the evaluation functional $f \to f(x)$ introduced in Theorem 34. Suppose that $X = \left\{x^{(k)}\right\}_{k=1}^{N}$ is a set of distinct points in $\mathbb{R}^d$. Then the reproducing kernel matrix is

$$R_{X,X} = \left(R_{x^{(j)}}(x^{(i)})\right).$$

I call this the reproducing kernel matrix because the reproducing kernel $K(x,y)$ satisfies $K(x,y) = R_x(y)$.

Using a separate symbol for the matrix $R_{X,X}$ fits in with the case of positive order, where $R_{X,X}$ is not simply a scalar multiple of $G_{X,X}$ but is related by a more complex formula.

The reproducing kernel matrix $R_{X,X}$ has the following properties:

Theorem 50 *The reproducing kernel matrix $R_{X,X}$ has the following properties:*

1. $R_{X,X} = (2\pi)^{-d/2} G_{X,X}$.

2. *$R_{X,X}$ is a Gram matrix and hence positive definite, Hermitian and regular.*

3. *The functions $\{R_{x^{(k)}}\}_{k=1}^N$ are independent.*

Proof. Part 1 True since in Theorem 34 it was shown that $R_x(z) = (2\pi)^{-d/2}\, G(z-x)$.

Part 2 $R_{X,X} = \left(R_{x^{(j)}}(x^{(i)})\right)$. But from the definition of R_x, $R_{x^{(j)}}(x^{(i)}) = (R_{x^{(j)}}, R_{x^{(i)}})_{w,0}$ so $R_{X,X}$ is a Gram matrix and hence is positive definite over $\mathbb{C}$, Hermitian and regular.

Part 3 In Theorem 45 it was shown that the functions $G(z-x^{(k)})$ are independent and since $R_x(z) = (2\pi)^{-d/2}\, G(z-x)$ the functions $R_{x^{(k)}}$ are independent. ■

2.4 The vector-valued evaluation operator $\widetilde{\mathcal{E}}_X$

The vector-valued evaluation operator $\widetilde{\mathcal{E}}_X$ and its Hilbert space adjoint $\widetilde{\mathcal{E}}_X^*$ are the fundamental operators used to solve the variational interpolation and smoothing problems.

Definition 51 ***The vector-valued evaluation operator*** $\widetilde{\mathcal{E}}_X$

Let $X = \left\{x^{(i)}\right\}_{i=1}^N$ be a set of N distinct points in $\mathbb{R}^d$. Let u be a complex-valued continuous function. Then the evaluation operator $\widetilde{\mathcal{E}}_X$ is defined by

$$\widetilde{\mathcal{E}}_X u = \left(u\left(x^{(i)}\right)\right)_{i=1}^N.$$

Sometimes we will use the notation u_X for $\widetilde{\mathcal{E}}_X u$ and when dealing with matrices $\widetilde{\mathcal{E}}_X u$ will be regarded as a column vector.

Theorem 52 *Let $X = \left\{x^{(i)}\right\}_{i=1}^N$ be a set of distinct points in $\mathbb{R}^d$ and suppose that the weight function w has property W2. Then we know that X_w^0 is a reproducing kernel Hilbert space of continuous functions and the evaluation functional $f \to f(x)$ has a Riesz representer, say R_x. The evaluation operator $\widetilde{\mathcal{E}}_X$ will be now shown to have the following properties:*

1. *$\widetilde{\mathcal{E}}_X : \left(X_w^0, \|\cdot\|_{w,0}\right) \to \left(\mathbb{C}^N, |\cdot|\right)$ is continuous, onto and* $\mathrm{null}\,\widetilde{\mathcal{E}}_X = W_{G,X}^{\perp}$.

2. *The adjoint operator $\widetilde{\mathcal{E}}_X^* : \mathbb{C}^N \to X_w^0$, defined by $\left(\widetilde{\mathcal{E}}_X f, g\right)_{\mathbb{C}^N} = \left(f, \widetilde{\mathcal{E}}_X^* g\right)_{w,0}$, satisfies*

$$\widetilde{\mathcal{E}}_X^* \beta = \sum_{i=1}^N \beta_i R_{x^{(i)}}, \quad \beta \in \mathbb{C}^N,$$

and is a homeomorphism from $\left(\mathbb{C}^N, |\cdot|\right)$ to $\left(W_{G,X}, \|\cdot\|_{w,0}\right)$.

3. *Suppose $R_{X,X} = \left(R_{x^{(j)}}\left(x^{(i)}\right)\right)$ is the reproducing kernel matrix and $\|\cdot\|$ is the matrix norm corresponding to the Euclidean vector norm. Then*

$$\left\|\widetilde{\mathcal{E}}_X^*\right\| = \left\|\widetilde{\mathcal{E}}_X\right\| = \|R_{X,X}\| \le \sqrt{N}\sqrt{R_0(0)}.$$

4. *The operator $\widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X$ is self-adjoint and we have $\left\|\widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X\right\| = \|R_{X,X}\|^2$, as well as the formulas*

$$\mathcal{E}_X^* \widetilde{\mathcal{E}}_X f = \sum_{i=1}^N f\left(x^{(i)}\right) R_{x^{(i)}}, \quad f \in X_w^0, \tag{2.2}$$

and

$$\left(\widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X f\right)(x) = \left(\widetilde{\mathcal{E}}_X f, \widetilde{\mathcal{E}}_X R_x\right)_{\mathbb{C}^N} = \left(f, \widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X R_x\right)_{w,0}, \quad f \in X_w^0. \tag{2.3}$$

Also, $\widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X : X_w^0 \to W_{G,X}$ is onto and $\mathrm{null}\,\widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X = W_{G,X}^{\perp}$.

5. *The operator* $\widetilde{\mathcal{E}}_X \widetilde{\mathcal{E}}_X^*$ *is self-adjoint and if we regard the range of* $\widetilde{\mathcal{E}}_X$ *as column vectors*
$$\widetilde{\mathcal{E}}_X \widetilde{\mathcal{E}}_X^* \beta = R_{X,X}\beta, \quad \beta \in \mathbb{C}^N,$$
where $R_{X,X}$ *is the (regular) reproducing kernel matrix introduced in Definition 49.*
6. *If* $X = \{x^{(i)}\}_{i=1}^N$ *and* $Y = \{y^{(j)}\}_{j=1}^{N'}$ *define* $R_{X,Y} = \left(R_{y^{(j)}}\left(x^{(i)}\right)\right)$ *and* $G_{X,Y} = \left(G\left(x^{(i)} - y^{(j)}\right)\right)$ *so that* $R_{X,Y} = (2\pi)^{-\frac{d}{2}} G_{X,Y}$. *Now when* $\widetilde{\mathcal{E}}_Y$ *is assumed to be a column vector we have* $\widetilde{\mathcal{E}}_X \widetilde{\mathcal{E}}_Y^* \beta = R_{X,Y}\beta$ *when* $\beta \in \mathbb{C}^{N'}$.

Proof. Parts 1 and 2 That $\widetilde{\mathcal{E}}_X$ is continuous follows from the Cauchy-Schwartz inequality and
$$\left|\widetilde{\mathcal{E}}_X u\right|_{\mathbb{C}^N}^2 = \sum_{i=1}^N \left|u(x^{(i)})\right|^2 = \sum_{i=1}^N \left|(u, R_{x^{(i)}})_{w,0}\right|^2 \le \left(\sum_{i=1}^N \|R_{x^{(i)}}\|_{w,0}^2\right) \|u\|_{w,0}^2 .$$
Clearly $\widetilde{\mathcal{E}}_X u = 0$ iff $(u, R_{x^{(i)}})_{w,0} = 0$ for each $R_{x^{(i)}}$ so that $\operatorname{null} \widetilde{\mathcal{E}}_X = W_{G,X}^{\perp}$.

Next we show that $\widetilde{\mathcal{E}}_X$ is onto. The Hilbert space adjoint of $\widetilde{\mathcal{E}}_X$ is denoted $\widetilde{\mathcal{E}}_X^*$ and is defined by
$$\left(\widetilde{\mathcal{E}}_X u, \beta\right)_{\mathbb{C}^N} = \left(u, \widetilde{\mathcal{E}}_X^* \beta\right)_{w,0} .$$
The adjoint is calculated using the representer R_x:
$$\left(\widetilde{\mathcal{E}}_X u, \beta\right)_{\mathbb{C}^N} = \sum_{i=1}^N u\left(x^{(i)}\right)\overline{\beta_i} = \sum_{i=1}^N (u, R_{x^{(i)}})_{w,0}\overline{\beta_i} = \left(u, \sum_{i=1}^N \beta_i R_{x^{(i)}}\right)_{w,0},$$
so that
$$\widetilde{\mathcal{E}}_X^* \beta = \sum_{i=1}^N \beta_i R_{x^{(i)}}, \quad \beta \in \mathbb{C}^N.$$
Finally we show that $\widetilde{\mathcal{E}}_X^*$ is 1-1. But $\widetilde{\mathcal{E}}_X^* \beta = 0$ implies $\sum_{i=1}^N \beta_i R_{x^{(i)}} = 0$ and, since $R_{x^{(i)}} = G\left(\cdot - x^{(i)}\right)$, Theorem 45 implies the $R_{x^{(i)}}$ are linearly independent and so $\beta = 0$.

Part 3 That $\left\|\widetilde{\mathcal{E}}_X^*\right\|_{op} = \left\|\widetilde{\mathcal{E}}_X\right\|_{op}$ is an elementary property of the adjoint. Now
$$\begin{aligned}\left\|\widetilde{\mathcal{E}}_X^*\beta\right\|_{w,0}^2 = \left(\sum_{i=1}^N \beta_i R_{x^{(i)}}, \sum_{j=1}^N \beta_j R_{x^{(j)}}\right)_{w,0} &= \sum_{i,j=1}^N \beta_i\overline{\beta_j}\left(R_{x^{(i)}}, R_{x^{(j)}}\right)_{w,0} \\ &= \sum_{i,j=1}^N \beta_i\overline{\beta_j} R_{x^{(j)}}\left(x^{(i)}\right) = \beta^T R_{X,X}\overline{\beta},\end{aligned}$$
so that, $\left\|\widetilde{\mathcal{E}}_X^*\right\| = \max\limits_{\beta\in\mathbb{C}^N} \frac{\|\widetilde{\mathcal{E}}_X^*\beta\|_{w,0}}{|\beta|} = \max\limits_{\beta\in\mathbb{C}^N} \frac{\sqrt{\beta^T R_{X,X}\overline{\beta}}}{|\beta|}$. But the latter expression is the largest (positive) eigenvalue of the Hermitian matrix $R_{X,X}$ i.e. the value of $\|R_{X,X}\|$. Hence $\left\|\widetilde{\mathcal{E}}_X^*\right\| = \|R_{X,X}\|$.
$$\begin{aligned}\left|\widetilde{\mathcal{E}}_X f\right|^2 = \sum_{k=1}^N \left|f\left(x^{(k)}\right)\right|^2 = \sum_{k=1}^N \left|(f, R_{x^{(k)}})_{w,0}\right|^2 &\le \sum_{k=1}^N \|f\|_{w,0}^2 \|R_{x^{(k)}}\|_{w,0}^2 \\ &= \sum_{k=1}^N \|f\|_{w,0}^2 R_0(0) = N R_0(0)\|f\|_{w,0}^2 .\end{aligned}$$

Part 4 The fact that $\widetilde{\mathcal{E}}_X^*\widetilde{\mathcal{E}}_X$ is self-adjoint and the formulas 2.2 and 2.3 are simple consequences of the definition of $\widetilde{\mathcal{E}}_X^*$ and R_x. That $\left\|\widetilde{\mathcal{E}}_X^*\widetilde{\mathcal{E}}_X\right\| = \left\|\widetilde{\mathcal{E}}_X\right\|^2$ is an elementary Hilbert space result and part 3 now implies $\left\|\widetilde{\mathcal{E}}_X^*\widetilde{\mathcal{E}}_X\right\| = \|R_{X,X}\|^2$.

From Definition 46 we know that the functions $\{R_{x^{(i)}}\}_{i=1}^N$ form a basis for $W_{G,X}$, where the $x^{(i)}$ are the unique points of X. Hence $\operatorname{range}\widetilde{\mathcal{E}}_X^* = W_{G,X}$ and $\operatorname{null}\widetilde{\mathcal{E}}_X^* = \{0\}$. Now recall the closed-range theorem, see for example Yosida [14], p205, which states that for a continuous linear operator $\mathcal{V}$, the range of $\mathcal{V}$ is closed iff the range of $\mathcal{V}^*$ is closed. Since $\operatorname{range}\widetilde{\mathcal{E}}_X^* = W_{G,X}$ and $W_{G,X}$ is finite dimensional, $\operatorname{range}\widetilde{\mathcal{E}}_X^*$ is closed and so $\operatorname{range}\widetilde{\mathcal{E}}_X$ is closed. Consequently, using the result that $\overline{\operatorname{range}\mathcal{V}} = (\operatorname{null}\mathcal{V}^*)^\perp$ for any continuous linear operator $\mathcal{V}$, it follows that

$$\operatorname{range}\widetilde{\mathcal{E}}_X = \overline{\operatorname{range}\widetilde{\mathcal{E}}_X} = \left(\operatorname{null}\widetilde{\mathcal{E}}_X^*\right)^\perp = \{0\}^\perp = \mathbb{C}^N.$$

Part 5

$$\widetilde{\mathcal{E}}_X\widetilde{\mathcal{E}}_X^*\beta = \widetilde{\mathcal{E}}_X\sum_{i=1}^N \beta_i R_{x^{(i)}} = \sum_{i=1}^N \widetilde{\mathcal{E}}_X R_{x^{(i)}}\beta_i = \left(\widetilde{\mathcal{E}}_X R_{x^{(1)}},\ldots,\widetilde{\mathcal{E}}_X R_{x^{(N)}}\right)\beta = R_{X,X}\beta,$$

since $R_{X,X} = \left(R_{x^{(j)}}\left(x^{(i)}\right)\right)$.

Part 6 The proof is very similar to that of part 5. ■

2.5 Minimal norm interpolation

2.5.1 The minimal norm interpolant and its solution

In this subsection we will formulate and solve the *minimal norm interpolation problem* for scattered data. The solution to this problem will be shown to be unique and its form derived.

Before defining the interpolation problem we will introduce some notation for the data. We assume the data is *scattered* i.e. irregularly spaced, as distinct from on a regular rectangular grid. The data is a set of distinct points $\left\{\left(x^{(i)}, y_i\right)\right\}_{i=1}^N$ with $x^{(i)} \in \mathbb{R}^d$ and $y_i \in \mathbb{C}$. We assume the $x^{(i)}$ must be distinct and set $X = \left\{x^{(i)}\right\}_{i=1}^N$ and $y = \{y_i\}_{i=1}^N$. In this document X will be called the *independent data* and y the *dependent data*.

The minimal norm interpolation problem
Suppose $\left\{\left(x^{(i)}, y_i\right)\right\}_{i=1}^N$ is scattered data, where the $x^{(i)}$ are distinct. We say $u_I \in X_w^0$ is a solution to the minimal norm interpolation problem if it interpolates the data and satisfies $\|u_I\|_{w,0} \le \|u\|_{w,0}$ for any other interpolant $u \in X_w^0$.

(2.4)

Using Hilbert space techniques we will now do the following:

1. Show there exists a unique minimal norm interpolant.
2. Show the interpolant is a basis function interpolant i.e. show that it lies in $W_{G,X}$.
3. Construct a matrix equation for the coefficients of the *data-translated* basis functions $G\left(\cdot - x^{(k)}\right)$.

The next theorem proves the minimal norm interpolation problem has a unique solution.

Theorem 53 *For given data there exists a unique minimal norm interpolant u_I. If v is any other interpolant satisfying $\widetilde{\mathcal{E}}_X v = y$ then*

$$\|u_I\|_{w,0}^2 + \|v - u_I\|_{w,0}^2 = \|v\|_{w,0}^2, \tag{2.5}$$

or equivalently

$$(v - u_I, u_I)_{w,0} = 0. \tag{2.6}$$

Proof. Since the evaluation operator $\widetilde{\mathcal{E}}_X : X_w^0 \to \mathbb{C}^N$ is continuous and onto, and the singleton set $\{y\}$ is closed, it follows that the set

$$\left\{u : \widetilde{\mathcal{E}}_X u = y\right\} = \widetilde{\mathcal{E}}_X^{-1}(y),$$

is a non-empty, proper, closed subspace of the Hilbert space X_w^0. Hence this subspace contains a unique element of smallest norm, say u_I. If $\widetilde{\mathcal{E}}_X v = y$ then

$$\left\{u : u \in X_w^0 \text{ and } \widetilde{\mathcal{E}}_X u = y\right\} = \left\{v - s : s \in \operatorname{null}\widetilde{\mathcal{E}}_X\right\}.$$

Now

$$\min\left\{\|v-s\|_{w,0} : s \in \operatorname{null}\widetilde{\mathcal{E}}_X\right\} = \operatorname{dist}\left(v, \operatorname{null}\widetilde{\mathcal{E}}_X\right),$$

is the distance between v and the closed subspace $\operatorname{null}\widetilde{\mathcal{E}}_X$. Therefore there exists a unique $s_I \in \operatorname{null}\widetilde{\mathcal{E}}_X$ such that

$$u_I = v - s_I, \tag{2.7}$$

$$\|v - s_I\|_{w,0} = \min\left\{\|v-s\|_{w,0} : s \in \operatorname{null}\widetilde{\mathcal{E}}_X\right\},$$

and

$$\|s_I\|_{w,0}^2 + \|v - s_I\|_{w,0}^2 = \|v\|_{w,0}^2 . \tag{2.8}$$

Substituting for s_I in 2.8 using 2.7 yields 2.5. Equation 2.6 follows since it is a necessary and sufficient condition for 2.5 to be true. ■

Next we prove that the minimal norm interpolant lies in the space $W_{G,X}$ of Definition 46.

Theorem 54 *For each data vector $y \in \mathbb{C}^N$ the minimal norm interpolant u_I is given by*

$$u_I = \widetilde{\mathcal{E}}_X^* R_{X,X}^{-1} y = (2\pi)^{d/2}\, \widetilde{\mathcal{E}}_X^* \left(G_{X,X}\right)^{-1} y, \tag{2.9}$$

where $\widetilde{\mathcal{E}}_X^* R_{X,X}^{-1} : \mathbb{C}^N \to W_{G,X}$ is an isomorphism. We also have

$$u_I(z) = \left(\widetilde{\mathcal{E}}_X \overline{R_z}\right)^T R_{X,X}^{-1} y. \tag{2.10}$$

Proof. From part 5 of Theorem 52, $\widetilde{\mathcal{E}}_X \widetilde{\mathcal{E}}_X^* = R_{X,X}$. Now by Theorem 50 $R_{X,X}$ is regular but in general $R_{X,X} \neq I$ so in general $\widetilde{\mathcal{E}}_X \widetilde{\mathcal{E}}_X^* y \neq y$ and $\widetilde{\mathcal{E}}_X^* y$ is not an interpolant. However, $\widetilde{\mathcal{E}}_X\left(\widetilde{\mathcal{E}}_X^* R_{X,X}^{-1}\right) = I$, so $\widetilde{\mathcal{E}}_X^* R_{X,X}^{-1} y$ is an interpolant in $W_{G,X}$. For convenience set $u = \widetilde{\mathcal{E}}_X^* R_{X,X}^{-1} y$. We want to show that $u = u_I$. But

$$(u, u_I - u)_{w,0} = \left(\widetilde{\mathcal{E}}_X^* R_{X,X}^{-1} y, u_I - u\right)_{w,0} = \left(R_{X,X}^{-1} y, \widetilde{\mathcal{E}}_X (u_I - u)\right)_{w,0} = 0,$$

since u_I and u are both interpolants to y. Thus

$$\|u_I\|_{w,0}^2 = \|u_I - u + u\|_{w,0}^2 = \|u_I - u\|_{w,0}^2 + \|u\|_{w,0}^2 ,$$

since $(u, u_I - u)_{w,0} = 0$. If $u \neq u_I$ then $\|u\|_{w,0} < \|u_I\|_{w,0}$, contradicting the fact that u_I is the minimal interpolant. Since $R_{X,X}$ is regular, $R_{X,X}^{-1}$ is an isomorphism from $\mathbb{C}^N$ to $\mathbb{C}^N$ and because $\widetilde{\mathcal{E}}_X^*$ is an isomorphism from $\mathbb{C}^N$ to $W_{G,X}$ we have that $\widetilde{\mathcal{E}}_X^* R_{X,X}^{-1} : \mathbb{C}^N \to W_{G,X}$ is an isomorphism.

Finally, using the definition of the adjoint $\widetilde{\mathcal{E}}_X^*$

$$u_I(z) = (u_I, R_z)_{w,0} = \left(\widetilde{\mathcal{E}}_X^* R_{X,X}^{-1} y, R_z\right)_{w,0} = \left(R_{X,X}^{-1} y, \widetilde{\mathcal{E}}_X R_z\right) = \left(\widetilde{\mathcal{E}}_X \overline{R_z}\right)^T R_{X,X}^{-1} y.$$

■

The last theorem allows us to define the mapping between a data function and its corresponding interpolant.

Definition 55 ***Data functions and the interpolant mapping*** $\mathcal{I}_X : X_w^0 \to W_{G,X}$
Given an independent data set X*, we shall assume that each member of* X_w^0 *can act as a legitimate data function* f *and generate the dependent data vector* $\widetilde{\mathcal{E}}_X f$*.*

Equation 2.9 of Theorem 54 enables us to define the linear mapping $\mathcal{I}_X : X_w^0 \to W_{G,X}$ *from the data functions to the corresponding unique minimal norm interpolant* $\mathcal{I}_X f = u_I$ *given by*

$$\mathcal{I}_X f = \widetilde{\mathcal{E}}_X^* R_{X,X}^{-1} \widetilde{\mathcal{E}}_X f, \quad f \in X_w^0. \tag{2.11}$$

Since $\widetilde{\mathcal{E}}_X \mathcal{I}_X f = \widetilde{\mathcal{E}}_X f$*, Theorem 52 can be easily used to show that* $\mathcal{I}_X$ *is a self-adjoint projection onto* $W_{G,X}$ *with null space* $W_{G,X}^\perp$*, and that* $\mathcal{I}_X f = f$ *iff* $f \in W_{G,X}$*. Since* f *interpolates the data it follows from Theorem 53 that*

$$\|f - \mathcal{I}_X f\|_{w,0} \leq \|f\|_{w,0}, \quad \|\mathcal{I}_X f\|_{w,0} \leq \|f\|_{w,0}, \quad f \in X_w^0, \tag{2.12}$$

i.e. $\mathcal{I}_X$ *and* $I - \mathcal{I}_X$ *are contractions.*

We now know that the interpolant lies in $W_{G,X}$ and from Theorem 45 we know that $\left\{G\left(\cdot - x^{(k)}\right)\right\}$ is a basis for $W_{G,X}$. The next step is to derive a matrix equation for the coefficients of the $G\left(\cdot - x^{(k)}\right)$:

Theorem 56 *The space* $W_{G,X}$ *contains only one interpolant to any given independent data vector* $y \in \mathbb{C}^N$*. This interpolant is the minimal norm interpolant and is defined uniquely by*

$$u_I(x) = \sum_{k=1}^{N} v_k G\left(x - x^{(k)}\right), \tag{2.13}$$

where the coefficient vector $v = (v_k)$ *satisfies the regular basis matrix equation*

$$G_{X,X} v = y. \tag{2.14}$$

Proof. From Theorem 54 we know that for a given independent data vector y, $W_{G,X}$ contains the minimal norm interpolant. Suppose $W_{G,X}$ contains another interpolant v_I. Then $\widetilde{\mathcal{E}}_X (u_I - v_I) = 0$ and so $u_I - v_I \in \operatorname{null} \widetilde{\mathcal{E}}_X$. But from part 1 of Theorem 52 $\operatorname{null} \widetilde{\mathcal{E}}_X = W_{G,X}^\perp$. Hence $u_I - v_I \in W_{G,X}^\perp$ and so $u_I = v_I$ since $W_{G,X}^\perp \cap W_{G,X} = \{0\}$.

Equation 2.13 and the interpolation requirement $u_I\left(x^{(k)}\right) = y_k$ implies $G_{X,X} v = y$ and by Theorem 50 $G_{X,X}$ is regular. ■

2.5.2 Error estimates derived without assuming unisolvent data sets

In this subsection we will prove several pointwise error estimates, also known as convergence estimates, for the minimal norm interpolant $\mathcal{I}_X f$ to its data function $f \in X_w^0$ as the independent data X 'fills' a bounded, closed data region K.

Unlike the results given by Light and Wayne for positive order in [6] and [7], the results of this subsection do not use Lagrange interpolation theory and so the proofs are much simpler. The order of convergence appears as the power of the radius of the largest ball that can be fitted between the data points X.

In this subsection we will consider two types of estimate which do not use unisolvent data sets. For want of better names we will just call them *Type 1* and *Type 2*. In the case of Type 1 the weight function has property W2 for some $\kappa \geq 0$ and the basis function G satisfies an extra smoothness condition near the origin. For *Type 2* estimates the weight function has property W2 for some $\kappa \geq 1$ and we use pointwise estimates based on the Riesz representer.

In the next subsection we will derive an error result which improves on the result of Type 2 by using the Lagrange interpolation theory. These results require that the independent data points constitute a unisolvent set and assume that the data region satisfy a cone condition. An order of convergence of $\lfloor \kappa \rfloor$ is then obtained.

When deriving error estimates we hope that as the number of evaluation points X increases i.e. as the 'minimum density' of the points increases, the interpolant will converge uniformly pointwise to the

data function f on the data region K. The minimum density of the independent data points is measured using the expression

$$h_{X,K} = \max_{x \in K} \operatorname{dist}(x, X) = \max_{x \in K} \min_{x^{(k)} \in X} \left| x - x^{(k)} \right|. \tag{2.15}$$

Here $h_{X,K}$ is the maximum distance between a point in K and a point in X, or equivalently $h_{X,K}$ is the radius of the largest open ball centered in K which does not contain any of the data points. Its use clearly makes sense intuitively but from a numerical point of view, in dimensions greater than 1, the calculation of $h_{X,K}$ presents formidable difficulties.

Theorem 34 tells us that the representer of the evaluation functional $f \to f(x)$ is $R_x = (2\pi)^{-d/2} G(\cdot - x) \in X_w^0$. Thus for any data function $f \in X_w^0$

$$f(x) - (\mathcal{I}_X f)(x) = (f - \mathcal{I}_X f, R_x)_{w,0},$$

where $\mathcal{I}_X f$ is the minimal interpolant of the data function on the independent data set X. Hence

$$\begin{aligned}|f(x) - (\mathcal{I}_X f)(x)| = \left| (f - \mathcal{I}_X f, R_x)_{w,0} \right| &\le \|f - \mathcal{I}_X f\|_{w,0} \|R_x\|_{w,0} \\ &= (2\pi)^{-d/4} \sqrt{G(0)} \|f - \mathcal{I}_X f\|_{w,0},\end{aligned} \tag{2.16}$$

where $G(0)$ is real and positive. The error information is clearly now restricted to the expression $\|f - \mathcal{I}_X f\|_{w,0}$ and if the inequality

$$\|f - \mathcal{I}_X f\|_{w,0} \le \|f\|_{w,0}, \tag{2.17}$$

implied by equation 2.5 is used the error information is completely lost. We can retain some of this information as follows. Choose any independent data point $x^{(j)}$. Then the interpolation property implies

$$f(x) - (\mathcal{I}_X f)(x) = (f - \mathcal{I}_X f)(x) - (f - \mathcal{I}_X f)\left(x^{(j)}\right). \tag{2.18}$$

In the results proved below we will use the estimates for $|f(x) - f(y)|$ and $|G(x) - G(y)|$ derived in Section 1.6. Before deriving the first error estimate I will show that there exists a suitable sequence of independent data sets $X^{(k)}$ with $h_{X^{(k)},K}$ tending to zero:

Theorem 57 *Suppose K is a bounded, closed, infinite set containing all the independent data sets. Then there exists a sequence of independent data sets $X^{(k)} \subset K$ such that $X^{(k)} \subset X^{(k+1)}$ and $h_{X^{(k)},K} \to 0$ as $k \to \infty$.*

Proof. For each $k = 1, 2, 3, \ldots$ there exists a finite covering of K by the balls $\left\{ B\left(a_k^{(j)}; \frac{1}{k}\right) \right\}_{j=1}^{M_k}$. Construct $X^{(1)}$ by choosing points from K so that one point lies in each ball $B\left(a_k^{(j)}; 1\right)$. Construct $X^{(k+1)}$ by first choosing the points $X^{(k)}$ and then at least one extra point so that $X^{(k+1)}$ contains points from each ball $B\left(a_{k+1}^{(j)}; \frac{1}{k+1}\right)$.

Then $x \in K \cap X^{(k)}$ implies $x \in B\left(a_k^{(j)}; \frac{1}{k}\right)$ for some j and hence $\operatorname{dist}\left(x, X^{(k)}\right) \le \frac{1}{k}$. Thus $h_{X^{(k)},K} = \max_{x \in K} \operatorname{dist}\left(x, X^{(k)}\right) \le \frac{1}{k}$ and $\lim_{k \to \infty} h_{X^{(k)},K} = 0$. ■

We start with some simple bounds for the pointwise error of the interpolant:

Theorem 58 *Suppose the weight function w has basis function G. Then the interpolant $\mathcal{I}_X f$ of any data function f satisfies*

$$|f(x) - (\mathcal{I}_X f)(x)| \le \sqrt{(f - \mathcal{I}_X f, f)_{w,0}} \sqrt{R_0(0)}, \quad x \in \mathbb{R}^d, \tag{2.19}$$

where $\sqrt{(f - \mathcal{I}_X f, f)_{w,0}} \le \|f\|_{w,0}$ and $R_0(0) = (2\pi)^{-\frac{d}{2}} G(0)$.

Also

$$|R_y(x) - (\mathcal{I}_X R_y)(x)| \le R_0(0), \quad x, y \in \mathbb{R}^d. \tag{2.20}$$

Proof. Since $\mathcal{I}_X$ is a self-adjoint projection w.r.t. $(\cdot,\cdot)_{w,0}$

$$\begin{aligned}\left|f(x)-(\mathcal{I}_X f)(x)\right| = \left|(f-\mathcal{I}_X f, R_x)_{w,0}\right| &\le \|f-\mathcal{I}_X f\|_{w,0}\|R_x\|_{w,0} \\ &= \sqrt{(f-\mathcal{I}_X f, f)_{w,0}}\sqrt{R_x(x)}.\end{aligned}$$

But $R_y(x) = (2\pi)^{-\frac{d}{2}} G(x-y)$ and inequalities 2.12 imply $\|f-\mathcal{I}_X f\|_{w,0} \le \|f\|_{w,0}$ so that $\sqrt{(f-\mathcal{I}_X f, f)_{w,0}} \le \|f\|_{w,0}$ and $R_x(x) = R_0(0) = (2\pi)^{-\frac{d}{2}} G(0)$. Now letting $f = R_y$ in 2.19 we get

$$\begin{aligned}\left|R_y(x)-(\mathcal{I}_X R_y)(x)\right| &\le \sqrt{(R_y-\mathcal{I}_X R_y, R_y)_{w,0}}\sqrt{R_0(0)} \\ &= \sqrt{R_y(y)-(\mathcal{I}_X R_y)(y)}\sqrt{R_0(0)},\end{aligned}$$

which implies that when $x=y$, $R_y(y)-(\mathcal{I}_X R_y)(y) \le R_0(0)$ and so

$$\left|R_y(x)-(\mathcal{I}_X R_y)(x)\right| \le R_0(0),$$

as required. ■

Type 1 pointwise estimates

Type 1 estimates place no *a priori* restriction on κ. In the next theorem a smoothness condition is applied to the basis function near the origin and this will allow a uniform order of convergence estimate to be obtained for the interpolant in a closed, bounded, infinite data region. The next result is based on part 2 of Corollary 36.

Theorem 59 *Suppose the weight function w has property W2 and that G is the basis function generated by w. Assume that for some $s>0$ and constants $C_G, h_G > 0$ the basis function satisfies*

$$G(0) - \operatorname{Re} G(x) \le C_G |x|^{2s}, \quad |x| < h_G. \tag{2.21}$$

Let $\mathcal{I}_X$ be the minimal norm interpolant mapping with the independent data set X contained in the closed, bounded, infinite set K, and let $k_G = (2\pi)^{-\frac{d}{4}}\sqrt{2C_G}$. Then for any data function $f \in X_w^0$ it follows that $\sqrt{(f-\mathcal{I}_X f, f)_{w,0}} \le \|f\|_{w,0}$ and

$$\left|f(x)-\mathcal{I}_X f(x)\right| \le k_G\sqrt{(f-\mathcal{I}_X f, f)_{w,0}}\,(h_{X,K})^s, \quad x\in K, \tag{2.22}$$

when $h_{X,K} = \max_{x\in K}\operatorname{dist}(x,X) < h_G$ i.e. the order of convergence is at least s.

Proof. From Theorem 58, $\sqrt{(f-\mathcal{I}_X f, f)_{w,0}} \le \|f\|_{w,0}$. Now fix $x\in K$ and let $X = \left\{x^{(j)}\right\}_{j=1}^N \subset K$ be an independent data set. Using the fact that $\mathcal{I}_X f$ interpolates f on X we can apply part 2 of Corollary 36 to obtain

$$\begin{aligned}\left|f(x)-\mathcal{I}_X f(x)\right| &= \left|(f-\mathcal{I}_X f)(x) - (f-\mathcal{I}_X f)\left(x^{(j)}\right)\right| \\ &\le (2\pi)^{-\frac{d}{4}}\sqrt{2}\,\|f-\mathcal{I}_X f\|_{w,0}\sqrt{G(0)-\operatorname{Re} G\left(x-x^{(j)}\right)} \\ &= (2\pi)^{-\frac{d}{4}}\sqrt{2}\sqrt{(f-\mathcal{I}_X f, f)_{w,0}}\sqrt{G(0)-\operatorname{Re} G\left(x-x^{(j)}\right)},\end{aligned}$$

for all j, where the last step used the fact that $\mathcal{I}_X$ is a self-adjoint projection. Then, noting that $\operatorname{dist}(x,X) < h_G$, we can apply the upper bound 2.21 and obtain

$$\begin{aligned}\left|f(x)-\mathcal{I}_X f(x)\right| &\le (2\pi)^{-\frac{d}{4}}\sqrt{2}\sqrt{(f-\mathcal{I}_X f, f)_{w,0}}\sqrt{C_G\left|x-x^{(j)}\right|^{2s}} \\ &= k_G\sqrt{(f-\mathcal{I}_X f, f)_{w,0}}\left|x-x^{(j)}\right|^s,\end{aligned}$$

for all j and so

$$\begin{aligned}\left|f(x)-\mathcal{I}_X f(x)\right| &\leq k_G \sqrt{(f-\mathcal{I}_X f, f)_{w,0}}\,(\operatorname{dist}(x,X))^s \\ &\leq k_G \sqrt{(f-\mathcal{I}_X f, f)_{w,0}}\left(\max_{x\in K}\operatorname{dist}(x,X)\right)^s \\ &= k_G \sqrt{(f-\mathcal{I}_X f, f)_{w,0}}\,(h_{X,K})^s,\end{aligned}$$

where the last step used inequalities 2.12. Since $\sqrt{(f-\mathcal{I}_X f, f)_{w,0}} \leq \|f\|_{w,0}$ the order of convergence is at least s. ■

We now prove a *double error rate* for the interpolant of data functions which are Riesz representers $R_{x'}$. This predicts a convergence order which is at least double that for an arbitrary data function. It also predicts a global bound for the error.

Corollary 60 *Under the notation and assumptions of the previous Theorem 59 the Riesz representer data functions $R_{x'}$ satisfy*

$$\left|R_{x'}(x)-(\mathcal{I}_X R_{x'})(x)\right| \leq (k_G)^2 (h_{X,K})^{2s}, \quad x\in K,\ x'\in\mathbb{R}^d,$$

when $h_{X,K}\leq h_G$ i.e. the order of convergence is at least $2s$ in $h_{X,K}$.

Proof. The key to this result is the term $\sqrt{(f, f-\mathcal{I}_X f)_{w,0}}$ in the estimate 2.22 proved in the last theorem. Substituting $f=R_{x'}$ in 2.22 gives

$$\begin{aligned}\left|R_{x'}(x)-(\mathcal{I}_X R_{x'})(x)\right| &\leq k_G \sqrt{(R_{x'}-\mathcal{I}_X R_{x'}, R_{x'})_{w,0}}\,(h_{X,K})^s \\ &= k_G \sqrt{R_{x'}(x')-(\mathcal{I}_X R_{x'})(x')}\,(h_{X,K})^s,\end{aligned}$$

so that

$$R_{x'}(x')-(\mathcal{I}_X R_{x'})(x') \leq k_G \sqrt{R_{x'}(x')-(\mathcal{I}_X R_{x'})(x')}\,(h_{X,K})^s,$$

which implies

$$\left|R_{x'}(x)-(\mathcal{I}_X R_{x'})(x)\right| \leq (k_G)^2 (h_{X,K})^{2s},$$

as required. ■

Examples: radial basis functions

For the shifted thin-plate splines, Gaussian, Sobolev and extended B-splines we will derive the interpolant error parameters s, h_G and C_G defined above in Theorem 59. To do this I will use the results of Theorem 40 when $\kappa\geq 1$: namely if w satisfies property W2 for $\kappa=1$ then

$$\left|G(0)-\operatorname{Re}G(x)\right| \leq C_G |x|^2, \quad x\in\mathbb{R}^d, \tag{2.23}$$

where

$$G(x)=f\left(r^2\right) \ \textit{implies}\ C_G=\|2rf''+f'\|_\infty, \tag{2.24}$$

$$G(x)=g(r) \quad \textit{implies}\ C_G=\frac{1}{2}\|g''\|_\infty. \tag{2.25}$$

If $\kappa<1$ I will use the mean value theorem.

Shifted thin-plate splines

From 1.4.2 the basis functions are given by

$$G(x)=(-1)^{\lceil v\rceil}\left(1+|x|^2\right)^v, \quad -d/2<v<0.$$

The weight function has property W2 for all $\kappa\geq 0$ so we can use the estimates 2.24 and 2.25. But G depends on $|x|^2$ so 2.24 is easier algebraically: $f(r)=(-1)^{\lceil v\rceil}(1+r)^v$, $f'(r)=(-1)^{\lceil v\rceil}v(1+r)^{v-1}$,

$f''(r) = (-1)^{\lceil v \rceil} v(v-1)(1+r)^{v-2}$ and $f'''(r) = (-1)^{\lceil v \rceil} v(v-1)(v-2)(1+r)^{v-3}$. For the function $|2rf''+f'|$ to take a maximum $3f''(r) = -2rf'''(r)$ must be satisfied i.e. $3v(v-1)(1+r)^{v-2} = -2v(v-1)(v-2)(1+r)^{v-3}$ which reduces to $3(1+r) = -2(v-2) = 2(2-v)$ so that

$$C_G = |(2rf''+f')(r_{\max})|, \quad \text{where } f(r) = (1+r)^v \text{ and } r_{\max} = \frac{(1-2v)}{3}. \tag{2.26}$$

Also $h_G = \infty$, $s = 1$ and $k_G = (2\pi)^{-\frac{d}{4}} \sqrt{2C_G}$. The order of convergence for an arbitrary data function is at least 1 and that for the Riesz representer is at least 2.

Gaussian

The weight function has property W2 for all $\kappa \geq 0$. It is easier algebraically to set $G(x) = f(r^2)$ where $f(r) = e^{-r}$ and use 2.24. For the function $2rf''+f'$ to take a maximum $3f''(r) = -2rf'''(r)$ must be satisfied i.e. $3e^{-r} = 2re^{-r}$ so that $r_{\max} = 3/2$ and hence $C_G = (2rf''+f')(r_{\max}) = (2re^{-r} - e^{-r})(r_{\max}) = 2e^{-3/2}$:

$$C_G = 2e^{-3/2}.$$

Also $h_G = \infty$, $s = 1$ and $k_G = (2\pi)^{-\frac{d}{4}} \sqrt{2C_G}$.

The order of convergence for an arbitrary data function is at least 1 and that for the Riesz representer is at least 2.

Sobolev splines

For this case the basis function is 1.15 i.e.

$$G(x) = \frac{1}{2^{v-1}\Gamma(v)} \widetilde{K}_{v-d/2}(|x|), \quad x \in \mathbb{R}^d, \tag{2.27}$$

constrained by $v > d/2$. The weight function has property W2 for $0 \leq \kappa < v - d/2$. There are two cases: $v - d/2$ is a positive integer and $v - d/2$ is a positive non-integer. Refer to Sections 3.1, 3.2 and 3.3 of Magnus et al. [9] for the properties of K_v and $\widetilde{K}_v$ given below: the smoothness results have been obtained from the infinite series representations.

$\boxed{\text{Case 1: } v - d/2 = 1, 2, 3, \ldots}$ Here we are dealing with the modified Bessel functions of integer order. Now rK_0 has properties

$$rK_0 \in C^{(0)}([0,\infty)) \cap C^{\infty}(0,\infty); \quad \lim_{r\to 0} rK_0(r) = 0; \quad K_0(r) > 0,\ r > 0;$$
$$\lim_{r\to\infty} rK_0(r) = 0.$$

and numerical experiments lead to the *hypothesis*: $rK_0(r)$ has a single turning (maximum) point:

$$0 \leq rK_0(r) \leq r'K_0(r') \lesssim 0.4665 \text{ when } 0 \leq r \leq r' \leq \arg\min \rho K_0(\rho) \simeq 0.595.$$

In general, for $m = 1, 2, 3, \ldots$, $\widetilde{K}_m = r^m K_m(r)$ satisfies

$$\widetilde{K}_{-m} = \widetilde{K}_m; \quad \widetilde{K}_m \in C^{(2m-1)}([0,\infty)) \cap C^{\infty}(0,\infty); \quad \widetilde{K}_m(0) = 2^{m-1}(m-1)!;$$
$$\widetilde{K}_m(r) > 0,\ r > 0,$$

the derivative of $\widetilde{K}_m$ satisfies

$$D\widetilde{K}_m(r) = -r\widetilde{K}_{m-1}(r); \quad D\widetilde{K}_m(0) = 0; \quad D\widetilde{K}_m(r) < 0, \quad x > 0;$$
$$\lim_{r\to\infty} D\widetilde{K}_m(r) = 0,$$

and if $m \geq 2$, the second derivative of $\widetilde{K}_m$ satisfies

$$D^2\widetilde{K}_m(r) = -\widetilde{K}_{m-1}(r) + r^2\widetilde{K}_{m-2}(r); \quad D^2\widetilde{K}_m(0) = -2^{m-2}(m-2)!\,; \tag{2.28}$$
$$\lim_{r\to\infty} D^2\widetilde{K}_m(r) = 0.$$

Now we want to estimate $|G(0) - \operatorname{Re} G(x)|$.

When $v - d/2 = 1$ we have $\kappa < 1$ and so we cannot use 2.23. However, by the mean value theorem

$$\left|\widetilde{K}_1(r) - \widetilde{K}_1(0)\right| \leq \max_{\rho\in[0,r]} \left|D\widetilde{K}_1(\rho)\right| r = \max_{\rho\in[0,r]} (\rho K_0(\rho))\, r \leq \|\rho K_0(\rho)\|_\infty r,$$

so that

$$\begin{aligned} G(0) - \operatorname{Re} G(x) &= \frac{1}{2^{v-1}\Gamma(v)} \left(\widetilde{K}_1(0) - \widetilde{K}_1(|x|)\right) \\ &\leq \frac{1}{2^{v-1}\Gamma(v)} \|\rho K_0(\rho)\|_\infty |x|, \end{aligned} \tag{2.29}$$

and consequently

$$if\ v - \frac{d}{2} = 1\ then\ s = \frac{1}{2},\ C_G = \frac{\|\rho K_0(\rho)\|_\infty}{2^{v-1}\Gamma(v)},\ h_G = \infty. \tag{2.30}$$

On the other hand if $v - d/2 = 2, 3, 4, \ldots$ then $\kappa \geq 1$ and we can use the estimate 2.23. Indeed, since

$$G(x) = g(r) = \frac{1}{2^{v-1}\Gamma(v)} \widetilde{K}_{n-d/2}(r), \tag{2.31}$$

$$C_G = \frac{1}{2}\|g''\|_\infty = \frac{1}{2^v\Gamma(v)} \left\|D^2\widetilde{K}_{v-d/2}\right\|_\infty, \quad v - \frac{d}{2} = 2, 3, 4, \ldots, \tag{2.32}$$

and thus

$$if\ v - \frac{d}{2} = 2, 3, 4, \ldots\ then\ s = 1,\ C_G\ is\ given\ by\ 2.32,\ h_G = \infty. \tag{2.33}$$

$\boxed{\text{Case 2: } v - d/2 > 0 \text{ and non-integer}}$ Here we are deal with the modified Bessel functions K_μ of positive non-integer order. We have

$$\begin{aligned} &\widetilde{K}_{-\mu} = \widetilde{K}_\mu; \quad \widetilde{K}_\mu \in C^\infty([0, \infty)); \quad \widetilde{K}_\mu(0) = 2^{|\mu|-1}\Gamma(|\mu|); \\ &\widetilde{K}_\mu(r) > 0,\ r > 0; \quad \lim_{r\to\infty} \widetilde{K}_\mu(r) = 0; \end{aligned}$$

Concerning the first derivatives:

$$D\widetilde{K}_\mu(r) = -r\widetilde{K}_{\mu-1}(r),$$

so that

$$D\widetilde{K}_\mu(0) = 0; \quad D\widetilde{K}_\mu(r) < 0, \quad r > 0; \quad \lim_{r\to\infty} D\widetilde{K}_\mu(r) = 0.$$

Regarding the second derivatives:

$$D^2\widetilde{K}_\mu(r) = -\widetilde{K}_{\mu-1}(r) + r^2\widetilde{K}_{\mu-2}(r),$$

so that

$$D^2\widetilde{K}_\mu(0) = -\widetilde{K}_{\mu-1}(0) = -2^{|\mu-2|}\Gamma(|\mu - 1|), \tag{2.34}$$

and

$$\lim_{r\to\infty} D^2\widetilde{K}_\mu(r) = 0.$$

Now we want to estimate $G(0) - \operatorname{Re} G(x)$.

When $0 < v - d/2 < 1$ we have $\kappa < 1$ and so we cannot use 2.23. However, using a second order Taylor series expansion (Appendix Subsection A.7)

$$\left|\widetilde{K}_\mu(r) - \widetilde{K}_\mu(0)\right| = \left|\widetilde{K}_\mu(r) - \widetilde{K}_\mu(0) - rD\widetilde{K}_\mu(0)\right| \leq \frac{r^2}{2} \max_{t\in[0,r]} \left|D^2\widetilde{K}_\mu(t)\right| \leq \frac{1}{2} \left\|D^2\widetilde{K}_\mu\right\|_\infty r^2,$$

so that

$$G(0) - \operatorname{Re} G(x) = \frac{1}{2^{v-1}\Gamma(v)} \left(\widetilde{K}_{v-d/2}(0) - \widetilde{K}_{v-d/2}(r)\right) \leq \frac{1}{2^v\Gamma(v)} \left\|D^2\widetilde{K}_{v-d/2}\right\|_\infty r^2,$$

and
$$C_G = \frac{1}{2^v \Gamma(v)} \left\| D^2 \widetilde{K}_{v-\frac{d}{2}} \right\|_\infty, \quad 0 < v - d/2 < 1. \tag{2.35}$$

When $v - d/2 > 1$, $v - \frac{d}{2} \notin \mathbb{Z}_+$ we have $\kappa \geq 1$ and so we can use 2.23 and 2.25 to get
$$C_G = \frac{1}{2} \|g''\|_\infty = \frac{1}{2^v \Gamma(v)} \left\| D^2 \widetilde{K}_{v-\frac{d}{2}} \right\|_\infty, \quad v - d/2 > 1,$$

and together with 2.35 we can conclude that:
$$s = 1,\ C_G \textit{ is given by } 2.35,\ h_G = \infty \text{ when } v - \frac{d}{2} \notin \mathbb{Z}_+,\ v - \frac{d}{2} > 0. \tag{2.36}$$

Examples: tensor product extended B-splines

With reference to the error estimate of Theorem 59:

Corollary 61 *Suppose the weight function is a tensor product extended B-spline weight function with parameters n and l. Then, if G_1 is the univariate basis function, the order of convergence s of the minimal norm interpolant to its data function is at least 1/2 and $C_G = \sqrt{d} G_1(0)^{d-1} \|DG_1\|_\infty$ and $h_G = \infty$.*

Proof. Theorem 43 showed that
$$|G(x) - G(y)| \leq \sqrt{d} G_1(0)^{d-1} \|DG_1\|_\infty |x - y|, \quad x, y \in \mathbb{R}^d,$$

so Theorem 59 gives this result. ■

Examples: summary table

Table 2.1 summarizes the convergence results 2.30, 2.33 and 2.36 for the Sobolev splines as predicted by Theorem 59, as well as those for the other Type 1 examples.

Interpolant error estimates **Type 1** non-unisolvent estimates: $\kappa \geq 0$, $k_G = (2\pi)^{-d/4} \sqrt{2C_G}$.				
Weight function	Parameter constraints	Converg. order s	C_G	h_G
Sobolev splines	$v - \frac{d}{2} = 1$	$\frac{1}{2}$	$\frac{\|\rho K_0(\rho)\|_\infty}{2^{v-1}\Gamma(v)}$	∞
$(v > d/2)$	$v - \frac{d}{2} \neq 1$	1	$\frac{\|D^2 K_{v-d/2}\|_\infty}{2^v \Gamma(v)}$	∞
Shifted thin-plate $(-d/2 < v < 0)$	-	1	eq. (2.26)	∞
Gaussian	-	1	$2e^{-3/2}$	∞
Extended B-spline	-	$\frac{1}{2}$	$G_1(0)^{d-1} \|DG_1\|_\infty \sqrt{d}$ (1)	∞

TABLE 2.1.

Another approach - the interpolation error seminorm

Here we will show how the definition of a seminorm using the interpolation operator can be used to obtain the above results concerning interpolation in a very concise manner. In Chapter 4 which studies a smoother (a stabilized interpolant) we will also make use of several seminorms to study the smoother error.

Definition 62 ***Interpolation error seminorm*** *and* ***semi-inner product***

Suppose we selected the semi-inner product $(f - \mathcal{I}_X f, g - \mathcal{I}_X g)_{w,0}$ for pointwise estimation of $f - \mathcal{I}_X f$. But $\mathcal{I}_X$ is a self-adjoint projection so $(f - \mathcal{I}_X f, g - \mathcal{I}_X g)_{w,0} = (f - \mathcal{I}_X f, g)_{w,0}$ and this is more suited to pointwise estimation since

$(f - \mathcal{I}_X f, R_x)_{w,0} = f(x) - (\mathcal{I}_X f)(x)$. So for $f, g \in X_w^0$ let us define the interpolation error seminorm and semi-inner product by
$$|f|_I = (f - \mathcal{I}_X f, f)_{w,0}, \quad \langle f, g \rangle_I = (f - \mathcal{I}_X f, g)_{w,0}.$$

The next theorem shows the close relationship between the pointwise error of the interpolant and the interpolation seminorm.

Theorem 63 *If $f,g \in X_w^0$ then some properties of the interpolation seminorm are:*

1. $\operatorname{null}|\cdot|_I = W_{G,X}$.
2. $|f|_I^2 = \|f\|_{w,0}^2 - \|\mathcal{I}_X f\|_{w,0}^2$ *and* $\langle f,g\rangle_I = (f,g)_{w,0} - (\mathcal{I}_X f, g)_{w,0}$.
3. $\langle f,g\rangle_I = (f,g)_{w,0} - \left(\widetilde{\mathcal{E}}_X f\right)^T (R_{X,X})^{-1} \widetilde{\mathcal{E}}_X g$.
4. $\langle f, R_x\rangle_I = f(x) - (\mathcal{I}_X f)(x)$.
5. $|R_x - R_{x^{(k)}}|_I = |R_x|_I$ *when* $x^{(k)} \in X$.

Proof. Using the comments following the definition of the data function and the interpolant map - Definition 55 - as a guide the proofs are straight forward. ■

The pointwise error of the interpolant has the following properties:

Theorem 64 *For all $x,y \in \mathbb{R}^d$:*

1. $|R_x(y) - (\mathcal{I}_X R_x)(y)|^2 \le (R_x(x) - (\mathcal{I}_X R_x)(x))(R_y(y) - (\mathcal{I}_X R_y)(y))$.
2. $R_x(x) - (\mathcal{I}_X R_x)(x) = \|R_x - R_{x^{(k)}}\|_{w,0}^2 - \|\mathcal{I}_X(R_x - R_{x^{(k)}})\|_{w,0}^2, \quad x^{(k)} \in X$.

Proof. Part 1. This is just the Cauchy-Schwartz inequality $\left|\langle R_x, R_y\rangle_I\right| \le |R_x|_I\,|R_y|_I$.

Part 2. From part 2 of Theorem 63
$|R_x - R_{x^{(k)}}|_I = \|R_x - R_{x_k}\|_{w,0}^2 - \|\mathcal{I}_X(R_x - R_{x_k})\|_{w,0}^2$ and from part 4 of Theorem 63 $|R_x|_I = R_x(x) - (\mathcal{I}_X R_x)(x)$. ■

From these two theorems and the calculation

$$\begin{aligned}
\|R_x - R_{x^{(k)}}\|_{w,0}^2 &= (R_x - R_{x^{(k)}}, R_x - R_{x^{(k)}})_{w,0} \\
&= R_x(x) - R_x\left(x^{(k)}\right) - \overline{R_x\left(x^{(k)}\right)} + R_{x^{(k)}}\left(x^{(k)}\right) \\
&= 2(2\pi)^{-d/2}\left(G(0) - \operatorname{Re} G\left(x - x^{(k)}\right)\right),
\end{aligned} \tag{2.37}$$

the interpolation results of Theorem 59 and Corollary 60 can be obtained without much difficulty.

Type 2 pointwise estimates

Type 2 pointwise estimates only assume $\kappa \ge 1$. The Type 1 estimates of Theorem 59 considered the case where the weight function had property W2 for some $\kappa \ge 0$ and an extra condition was applied to the basis function. In the next theorem we only assume that $\kappa \ge 1$ and derive an order of convergence estimate of 1, as well as a doubled convergence estimate of 2 for the Riesz data functions R_y.

Substituting $\alpha = \beta = 0$ and $y = 0$ in part 3 of Corollary 37 gives the Riesz representer-type estimate

$$G(0) - \operatorname{Re} G(x) \le \frac{1}{(2\pi)^{\frac{d}{2}}}\left(\int \frac{d\xi}{w(\xi)}\right)^{\frac{1}{2}}\left(\int \frac{|\xi|^2\, d\xi}{w(\xi)}\right)^{\frac{1}{2}} |x|,$$

since $G(0)$ is real and $|G(x)| \le G(0)$. Hence G always satisfies 2.21 with $s = 1/2$ and the theory of Type 1 convergence estimates can be always be applied to obtain convergence rates of at least 1/2. Also, $\kappa \ge 1$ implies $G \in C^{(\lfloor 2\kappa \rfloor)} \subseteq C^{(2)}$ and the mean value theorem or the Taylor series expansion can always be applied with $s = 1/2$ or $s = 1$ to obtain Type 1 convergence estimates.

However, instead we will start by using part 1 of Corollary 36 - a Riesz representer-type estimate for data functions - to obtain some error estimates. Note that the constant k_G in the next Theorem is defined to match the constant k_G for the Type 1 convergence estimate 2.22.

Theorem 65 *Suppose the weight function w has property W2 for a parameter $\kappa \geq 1$. Suppose also that $\mathcal{I}_X f$ is the minimal interpolant on X of the data function $f \in X_w^0$ and that K is a bounded, closed, infinite subset of $\mathbb{R}^d$, $X \subset K$ and $h_{X,K} = \max_{x \in K} \operatorname{dist}(x, X)$.*

Then

$$|f(x) - (\mathcal{I}_X f)(x)| \leq k_G \sqrt{(f - \mathcal{I}_X f, f)_{w,0}}\, h_{X,K}, \quad x \in K, \tag{2.38}$$

where

$$k_G^2 = (2\pi)^{-d} \int \frac{|\cdot|^2}{w} = -(2\pi)^{-\frac{d}{2}} \left(\nabla^2 G\right)(0), \quad \sqrt{(f - \mathcal{I}_X f, f)_{w,0}} \leq \|f\|_{w,0}, \tag{2.39}$$

and the order of convergence is at least 1. Further, for the 'Riesz' data functions R_y, $y \in K$, we have

$$|R_y(x) - (\mathcal{I}_X R_y)(x)| \leq (k_G)^2 (h_{X,K})^2, \quad x, y \in K, \tag{2.40}$$

i.e. a 'doubled' estimate of at least 2 for the rate of convergence.

Proof. Let $X = \left\{x^{(j)}\right\}_{j=1}^N$. Then since $\mathcal{I}_X$ is an interpolant on X

$$f(x) - (\mathcal{I}_X f)(x) = (f - \mathcal{I}_X f)(x) - (f - \mathcal{I}_X f)\left(x^{(j)}\right),$$

for all j and applying the inequality of part 1 of Corollary 36 to $f - \mathcal{I}_X f \in X_w^0$ we get

$$\begin{aligned} |f(x) - (\mathcal{I}_X f)(x)| &= \left|(f - \mathcal{I}_X f)(x) - (f - \mathcal{I}_X f)\left(x^{(j)}\right)\right| \\ &\leq k_G \|f - \mathcal{I}_X f\|_{w,0} \left|x - x^{(j)}\right| \\ &= k_G \sqrt{(f - \mathcal{I}_X f, f)_{w,0}} \left|x - x^{(j)}\right|, \quad j = 1, \ldots, N, \end{aligned} \tag{2.41}$$

Now suppose that $X \subset K$. Then the inequalities 2.41 imply that

$$|f(x) - (\mathcal{I}_X f)(x)| \leq k_G \sqrt{(f - \mathcal{I}_X f, f)_{w,0}} \operatorname{dist}(x, X),$$

and so if x is restricted to K

$$|f(x) - (\mathcal{I}_X f)(x)| \leq k_G \sqrt{(f - \mathcal{I}_X f, f)_{w,0}} \max_{x \in K} \operatorname{dist}(x, X) = k_G \sqrt{(f - \mathcal{I}_X f, f)_{w,0}}\, h_{X,K}.$$

Thus when $f = R_y$

$$\begin{aligned} |R_y(x) - (\mathcal{I}_X R_y)(x)| &\leq k_G \sqrt{(R_y - \mathcal{I}_X R_y, R_y)_{w,0}}\, h_{X,K} \\ &= k_G \sqrt{R_y(y) - (\mathcal{I}_X R_y)(y)}\, h_{X,K}, \end{aligned} \tag{2.42}$$

so that when $x = y$

$$R_y(y) - (\mathcal{I}_X R_y)(y) \leq k_G \sqrt{R_y(y) - (\mathcal{I}_X R_y)(y)}\, h_{X,K},$$

and $R_y(y) - (\mathcal{I}_X R_y)(y) \leq (k_G h_{X,K})^2$. Inequality 2.42 now implies inequality 2.40 as required.

Finally, that $\int \frac{|\cdot|^2}{w} = -(2\pi)^{d/2} \left(\nabla^2 G\right)(0)$, follows directly from 1.45. ■

Examples: radial basis functions

It was shown in Subsection 1.2.3 that the **shifted thin-plate splines** and the **Gaussian** have weight functions satisfying property W2 for all $\kappa \geq 1$. Thus we can apply Theorem 65 to obtain the same orders of convergence that were obtained for the Type 1 results in Subsubsection 2.5.2 i.e. a convergence of order 1.

It was shown in Subsection 1.2.3 that $0 \leq \kappa < v - d/2$ for the **Sobolev splines**. Thus we can choose $\kappa \geq 1$ iff $v - d/2 > 1$.

From 2.39, $k_G^2 = -(2\pi)^{-d}\left(\nabla^2 G\right)(0)$. Now if $G(x) = f\left(r^2\right)$ then by part 2 of Theorem 39, $\left(\nabla^2 G\right)(0) = 2f'(0)\,d$ and so

$$k_G = (2\pi)^{-d/4}\sqrt{-2f'(0)}\sqrt{d}, \tag{2.43}$$

and if $G(x) = g(r)$ then by part 1 of Theorem 39, $\left(\nabla^2 G\right)(0) = g''(0)\,d$ and so

$$k_G = (2\pi)^{-d/4}\sqrt{-g''(0)}\sqrt{d}. \tag{2.44}$$

Shifted thin-plate splines $f(r) = (1+r)^v$, $f\prime(r) = v(1+r)^{v-1}$ and so $k_G = (2\pi)^{-d/4}\sqrt{-2v}\sqrt{d}$.

Gaussian $f(r) = e^{-r}$, $f\prime(r) = -e^{-r}$ and hence $k_G = (2\pi)^{-d/4}\sqrt{2}\sqrt{d}$.

Sobolev splines Here we use results of Subsubsubsection 2.5.2. From 2.27,
$g(r) = \frac{1}{2^{v-1}\Gamma(v)}\widetilde{K}_{v-d/2}(r)$ and so $g''(0) = \frac{1}{2^{v-1}\Gamma(v)}D^2\widetilde{K}_{v-d/2}(0)$.

<u>If $v-d/2$ is an integer</u> then $v-d/2 = 2,3,4,\ldots$ and by 2.28

$$g''(0) = \frac{D^2\widetilde{K}_{v-d/2}(0)}{2^{v-1}\Gamma(v)} = -\frac{2^{v-d/2-2}(v-d/2-2)!}{2^{v-1}\Gamma(v)} = -\frac{(v-d/2-2)!}{2^{d/2+1}\Gamma(v)},$$

and by 2.44

$$k_G = (2\pi)^{-d/4}\sqrt{-g''(0)}\sqrt{d} = (2\pi)^{-d/4}\sqrt{\frac{(v-d/2-2)!}{2^{d/2+1}\Gamma(v)}}\sqrt{d}. \tag{2.45}$$

<u>If $v-d/2$ is **not** an integer</u> then $v-d/2 > 1$ and $v-d/2 \neq 2,3,4,\ldots$. By 2.34

$$\begin{aligned} g''(0) = \frac{D^2\widetilde{K}_{v-d/2}(0)}{2^{v-1}\Gamma(v)} &= -\frac{2^{|v-d/2-2|}\Gamma(|v-d/2-1|)}{2^{v-1}\Gamma(v)} = -\frac{2^{|v-d/2-2|}\Gamma(v-d/2-1)}{2^{v-1}\Gamma(v)} \\ &= -\frac{\Gamma(v-d/2-1)}{2^{v-1-|v-d/2-2|}\Gamma(v)}, \end{aligned}$$

and

$$\begin{aligned} k_G &= (2\pi)^{-d/4}\sqrt{\frac{\Gamma(v-d/2-1)}{2^{v-1-|v-d/2-2|}\Gamma(v)}}\sqrt{d} \\ &= \begin{cases} (2\pi)^{-d/4}\sqrt{\frac{\Gamma(v-d/2-1)}{2^{2v-d/2-3}\Gamma(v)}}\sqrt{d}, & 1 < v-d/2 < 2, \\ (2\pi)^{-d/4}\sqrt{\frac{\Gamma(v-d/2-1)}{2^{d/2+1}\Gamma(v)}}\sqrt{d}, & v-d/2 > 2, \\ & v-d/2 \neq 2,3,4,\ldots \end{cases} \end{aligned} \tag{2.46}$$

Finally, 2.45 and 2.46 can be combined to give

$$k_G = \begin{cases} (2\pi)^{-d/4}\sqrt{\frac{\Gamma(v-d/2-1)}{2^{2v-d/2-3}\Gamma(v)}}\sqrt{d}, & 1 < v-d/2 < 2, \\ (2\pi)^{-d/4}\sqrt{\frac{\Gamma(v-d/2-1)}{2^{d/2+1}\Gamma(v)}}\sqrt{d}, & v-d/2 > 2. \end{cases}$$

<u>Examples: tensor product basis functions</u>

Extended B-splines $(1 \leq n \leq l)$ Since the weight function has property W2 for κ iff $\kappa + 1/2 < n$ we can choose $\kappa = 1$ iff $n \geq 2$. Then by Theorem 65, $k_G = (2\pi)^{-d/4}\sqrt{-\left(\nabla^2 G\right)(0)}$ and since

$$\begin{aligned} \left(\nabla^2 G\right)(0) = \sum_{k=1}^{d}\left(D_k^2 G\right)(0) &= \sum_{k=1}^{d} D_k^2\left(G_1(x_1)\,G_1(x_2)\ldots G_1(x_d)\right)(0) \\ &= \sum_{k=1}^{d} G_1(0)^{d-1}D^2G_1(0) \\ &= G_1(0)^{d-1}D^2G_1(0)\,d, \end{aligned}$$

it follows that

$$k_G = (2\pi)^{-d/4} \sqrt{-G_1(0)^{d-1} D^2 G_1(0)} \sqrt{d}.$$

These convergence results are summarized in the Table 2.2.

Examples: summary table

Interpolant convergence order estimates **Type 2** non-unisolvent estimates: $\kappa \geq 1$				
Weight function	Parameter constraints	Converg. order	$(2\pi)^{d/4} k_G$	h_G
Sobolev splines	$v - \frac{d}{2} \geq 2$	1	$\sqrt{\frac{\Gamma(v-d/2-1)}{2^{d/2+1}\Gamma(v)}}\sqrt{d}$	∞
$(v > d/2)$	$1 < v - \frac{d}{2} < 2$	1	$\sqrt{\frac{\Gamma(v-d/2-1)}{2^{2v-d/2-3}\Gamma(v)}}\sqrt{d}$	∞
Shifted thin-plate $(-d/2 < v < 0)$	-	1	$\sqrt{-2v}\sqrt{d}$	∞
Gaussian	-	1	$\sqrt{2}\sqrt{d}$	∞
Extended B-spline $(1 \leq n \leq l)$	$n \geq 2$	1	$\sqrt{-G_1(0)^{d-1} D^2 G_1(0)}\sqrt{d}$	∞

TABLE 2.2.

2.5.3 Pointwise error estimates using unisolvent data sets when $\kappa \geq 1$

Unisolvent sets of independent data points are fundamental to the theory of positive order basis function interpolation. In that case the order of the unisolvent set is the order of the basis function. We will now show that this technique can be used when the order is zero. Here, however, the order of unisolvency is related to the value of the weight function parameter κ.

In Theorem 65 we assumed the weight function had property W2 for some $\kappa \geq 1$ and so obtained an order 1 estimate for the convergence of the interpolant. In this subsection we will show that by assuming the data is a unisolvent set of points of order $\lfloor \kappa \rfloor$ an order $\lfloor \kappa \rfloor$ convergence estimate can be obtained for and arbitrary data function.

Definition 66 ***Unisolvent sets and minimal unisolvent sets***
Recall that P_m is the set of polynomials of order m i.e. of degree $m-1$ when $m \geq 1$.

Then a finite set of distinct points $X = \{x_i\}$ is said to be a ***unisolvent set*** *with respect to P_m if: $p \in P_m$ and $p(x_i) = 0$ for all $x_i \in X$ implies $p = 0$.*

*Sometimes we say X is unisolvent of order m or that X is m-**unisolvent**.*

It is known that any unisolvent set has at least $M = \dim P_m$ points, and that any unisolvent set of more than M points has a unisolvent subset with M points. Consequently, a unisolvent set with M points is called a ***minimal unisolvent set****.*

Theorem 67 *Suppose $M = \dim P_m$ and $A = \{a_i\}_{i=1}^M \subset \mathbb{R}^d$.*

1. *If $\{p_j\}_{j=1}^M$ is any basis of P_m then A is minimally m-unisolvent iff $\det(p_j(a_i)) \neq 0$.*
2. *In one dimension any set of m distinct points is m-unisolvent.*

Proof. Part 1 follows directly from the definition of unisolvency. **Part 2** Choose the monomial basis $\{x^k\}_{k=0}^{m-1}$ for P_m. Then the determinant of part 1 is Vandermonde's alternant which is singular iff two points coincide. ■

Definition 68 ***Cardinal basis for polynomials*** P_m.
A basis $\{l_i\}_{i=1}^M$ for P_m is a ***cardinal basis*** *for the minimal unisolvent set $A = \{a_i\}_{i=1}^M$ if $l_i(a_j) = \delta_{i,j}$ and the l_i are polynomials with* ***real valued*** *coefficients.*

It is known that a set is a minimal unisolvent iff there exists a (unique) cardinal basis for the set.

The next step is to introduce the Lagrange interpolation operator $\mathcal{P}$ and the operator $\mathcal{Q} = I - \mathcal{P}$:

Definition 69 ***The operators*** $\mathcal{P} : C^{(0)} \to P_m$ ***and*** $\mathcal{Q} : C^{(0)} \to C^{(0)}$.

These operators are only defined for integers $m \geq 1$. *Suppose the set* $A = \{a_i\}_{i=1}^M$ *is a minimal unisolvent set with respect to the polynomials* P_m *and by Definition 68 there is a unique cardinal basis* $\{l_i\}_{i=1}^M$ *for* A. *Then for any continuous function* f *the operators* $\mathcal{P}$ *and* $\mathcal{Q}$ *are defined by*

$$\mathcal{P}f = \sum_{i=1}^M f(a_i)\, l_i, \qquad \mathcal{Q}f = f - \mathcal{P}f.$$

Theorem 70 ***The operators*** $\mathcal{P}$ ***and*** $\mathcal{Q}$ *have the following elementary properties:*

1. $p \in P_m$ *implies* $\mathcal{P}p = p$ *and hence* $\mathcal{Q}p = 0$.
2. $\mathcal{P}f$ *interpolates the data* $\{\ (a_i, f(a_i))\}_{i=1}^M$. *Indeed,* $\mathcal{P}$ *is termed the* ***Lagrange polynomial interpolation function***.
3. $\mathcal{P}^2 = \mathcal{P}$, $\mathcal{P}\mathcal{Q} = \mathcal{Q}\mathcal{P} = 0$ *and* $\mathcal{Q}^2 = \mathcal{Q}$ *so the operators are projections.*
4. $\mathrm{null}\, \mathcal{Q} = \mathrm{range}\, \mathcal{P}$.

Proof. Part 1 is true since each member of the cardinal basis satisfies $\mathcal{P}l_j = \sum\limits_{i=1}^M l_i(a_i)\, l_i = l_j$. **Part 2** is true because $l_i(a_j) = \delta_{i,j}$. Regarding **part 3**, $\mathcal{P}$ is a projection since by part 2

$$\mathcal{P}^2 f = \mathcal{P}\left(\sum_{i=1}^M f(a_i)\, l_i\right) = \sum_{|\alpha|<k} f(a_i)\, \mathcal{P}(l_i) = \sum_{i=1}^M f(a_i)\, l_i = \mathcal{P}f,$$

and so

$$\mathcal{P}\mathcal{Q} = \mathcal{P}(I - \mathcal{P}) = \mathcal{P} - \mathcal{P}^2 = 0 = (I - \mathcal{P})\,\mathcal{P} = \mathcal{Q}\mathcal{P},$$

which implies $\mathcal{Q}^2 = \mathcal{Q}(I - \mathcal{P}) = \mathcal{Q}$ and we have proved part 3. Finally $\mathcal{P} + \mathcal{Q} = I$ so **part 4** holds. ■

To study the pointwise order of convergence of the minimal interpolant to its data function we will need the following two lemmas. The first lemma provides a pointwise upper bound for the operator $\mathcal{Q}$ using a *multipoint Taylor series expansion* which will express $\mathcal{Q}f(x) = u(x) - \mathcal{P}u(x)$ in terms of the derivatives of order m evaluated on the intervals $[x, a_i]$. The basic tool is the Taylor series expansion with integral remainder.

Lemma 71 *Suppose* $f \in C^{(m)}(\mathbb{R}^d)$ *and* $A = \{a_i\}_{i=1}^M$ *is a minimal unisolvent set of order* $m \geq 1$. *Then we have the upper bound*

$$|\mathcal{Q}f(x)| \leq \frac{d^{\frac{m}{2}}}{m!}\left(\sum_{i=1}^M |l_i(x)|\right)\left(\max_{\substack{|\beta|=m \\ z\in S_{A,x}}} \left|\left(D^\beta f\right)(z)\right|\right)\left(\max_{i=1}^M |a_i - x|\right)^m, \quad x \in \mathbb{R}^d,$$

where $S_{A,x}$ *is the union of closed intervals* $S_{A,x} = \bigcup\limits_{i=1}^M [x, a_i]$.

Proof. From Definition 69 of $\mathcal{P}$ and $\mathcal{Q}$

$$\mathcal{Q}f(x) = f(x) - \mathcal{P}f(x) = f(x) - \sum_{i=1}^M f(a_i)\, l_i(x) = f(x) - \sum_{i=1}^M f(x + (a_i - x))\, l_i(x).$$

Using the Taylor series expansion formula with integral remainder given in Subsection A.7 of the Appendix, we have for each i

$$f(x + (a_i - x)) = \sum_{|\beta|<m} \frac{D^\beta f(x)}{\beta!}(a_i - x)^\beta + (\mathcal{R}_m f)(x, a_i - x),$$

where the remainder term satisfies

$$\left|(\mathcal{R}_m f)(x, a_i - x)\right| \leq \frac{d^{\frac{m}{2}}}{m!} \max_{\substack{|\beta|=m \\ t\in[x,a_i]}} \left|D^\beta f(t)\right| \; |a_i - x|^m .$$

Thus

$$\begin{aligned}\sum_{i=1}^{M} f(x + (a_i - x))\, l_i(x) &= \sum_{i=1}^{M} \left(\sum_{|\beta|<m} \frac{D^\beta f(x)}{\beta!} (a_i - x)^\beta + (\mathcal{R}_m f)(x, a_i - x) \right) l_i(x) \\ &= \sum_{i=1}^{M} \sum_{|\beta|<m} \frac{D^\beta f(x)}{\beta!} (a_i - x)^\beta\, l_i(x) + \sum_{i=1}^{M} (\mathcal{R}_m f)(x, a_i - x)\, l_i(x) .\end{aligned}$$

But by part 1 of Theorem 70 the operator $\mathcal{P}$ preserves polynomials of degree $< m$. Hence

$$\begin{aligned}\sum_{i=1}^{M} \sum_{|\beta|<m} \frac{D^\beta f(x)}{\beta!} (a_i - x)^\beta\, l_i(x) &= \sum_{|\beta|<m} \frac{D^\beta f(x)}{\beta!} \mathcal{P}_y \left((y - x)^\beta \right)(y = x) \\ &= \sum_{|\beta|<m} \frac{D^\beta f(x)}{\beta!} \left((y - x)^\beta \right)(y = x) \\ &= f(x) ,\end{aligned}$$

leaving us with

$$\sum_{i=1}^{M} f(x + (a_i - x))\, l_i(x) = f(x) + \sum_{i=1}^{M} (\mathcal{R}_m f)(x, a_i - x)\, l_i(x) ,$$

and

$$\mathcal{Q} f(x) = -\sum_{i=1}^{M} (\mathcal{R}_m f)(x, a_i - x)\, l_i(x) .$$

Finally, by applying the remainder estimate A.3 given in the Appendix we get

$$\begin{aligned}|\mathcal{Q} f(x)| &\leq \left(\sum_{i=1}^{M} |l_i(x)| \right) \max_{i=1}^{M} |(\mathcal{R}_m f)(x, a_i - x)| \\ &\leq \frac{d^{\frac{m}{2}}}{m!} \left(\sum_{i=1}^{M} |l_i(x)| \right) \max_{i=1}^{M} \left(\max_{\substack{|\beta|=m \\ t\in[x,a_i]}} \left|D^\beta f(t)\right| \; |a_i - x|^m \right) \\ &\leq \frac{d^{\frac{m}{2}}}{m!} \left(\sum_{i=1}^{M} |l_i(x)| \right) \left(\max_{\substack{|\beta|=m \\ z\in S_{A,x}}} \left|D^\beta f(z)\right| \right) \left(\max_{i=1}^{M} |a_i - x| \right)^m ,\end{aligned}$$

as required. ■

To study the convergence of the minimal interpolant when the data is unisolvent we will also need the following lemma which supplies the required results from the theory of Lagrange interpolation. These results are stated without proof. This lemma has been created from Lemma 3.2, Lemma 3.5 and Theorem 3.6 of Light and Wayne [7]. The results of this lemma do not involve any reference to weight or basis functions or to functions in X_w^0, but use the properties of the data region Ω which contains the independent data points X and the order of the unisolvency used for the interpolation. Thus we have separated the part of the proof that involves basis functions from the part that uses the detailed theory of Lagrange interpolation operators.

Lemma 72 *Suppose first that:*

1. *Ω is a bounded, open, connected subset of $\mathbb{R}^d$ having the cone property e.g. §4.3 Adams [1].*

2. X is a unisolvent subset of Ω of order m.

Suppose $\{l_j\}_{j=1}^{M}$ is the cardinal basis of P_m with respect to a minimal unisolvent subset of Ω. Using Lagrange polynomial interpolation techniques, it can be shown there exists a constant $K'_{\Omega,m} > 0$ such that

$$\sum_{j=1}^{M} |l_j(x)| \leq K'_{\Omega,m}, \quad x \in \overline{\Omega}, \tag{2.47}$$

and all minimal unisolvent subsets of Ω. Now define

$$h_{X,\Omega} = \sup_{\omega \in \Omega} \operatorname{dist}(\omega, X),$$

and fix $x \in X$. By using Lagrange interpolation techniques it can be shown there are constants $c_{\Omega,m}, h_{\Omega,m} > 0$ such that when $h_{X,\Omega} < h_{\Omega,m}$ there exists a minimal unisolvent set $A \subset X$ satisfying

$$\operatorname{diam}(A \cup \{x\}) \leq c_{\Omega,m} h_{X,\Omega}.$$

Before deriving the first interpolation error estimate I will show that there exists a suitable sequence of independent data sets $X^{(k)} \subset \Omega$ with $h_{X^{(k)},\Omega}$ tending to zero. The proof is very close to that of Theorem 57 which considered the case of a closed data region.

Theorem 73 *Suppose Ω is a bounded, open set containing all the independent data sets. Then there exists a sequence of independent data sets $X^{(k)} \subset \Omega$ such that $X^{(k)} \subset X^{(k+1)}$ and $h_{X^{(k)},\Omega} \to 0$ as $k \to \infty$.*

Proof. For $k = 1, 2, 3, \ldots$ there exists a finite covering of Ω by the balls $\left\{B\left(a_k^{(j)}; \frac{1}{k}\right)\right\}_{j=1}^{M_k}$. Construct $X^{(1)}$ by choosing points from Ω so that one point lies in each ball $B\left(a_k^{(j)}; 1\right)$. Construct $X^{(k+1)}$ by first choosing the points $X^{(k)}$ and then at least one extra point so that $X^{(k+1)}$ contains points from each ball $B\left(a_{k+1}^{(j)}; \frac{1}{k+1}\right)$.

Then $x \in \Omega \cap X^{(k)}$ implies $x \in B\left(a_k^{(j)}; \frac{1}{k}\right)$ for some j and hence $\operatorname{dist}\left(x, X^{(k)}\right) < \frac{1}{k}$. Hence $h_{X^{(k)},\Omega} = \sup_{x \in \Omega} \operatorname{dist}\left(x, X^{(k)}\right) < \frac{1}{k}$ and $\lim_{k \to \infty} h_{X^{(k)},\Omega} = 0$. ■

Now we are ready to state our order of convergence result for the minimal norm interpolant for the case of a weight function with integer parameter $\kappa \geq 1$.

Theorem 74 *Let w be a weight function with property W2 for integer parameter $\kappa \geq 1$ and let G be the corresponding basis function. Set $m = \lfloor \kappa \rfloor$.*

Suppose $\mathcal{I}_X f$ is the minimal norm interpolant of the data function $f \in X_w^0$ on the independent data set X contained in the data region Ω.

We use the notation and assumptions of Lemma 72 which means assuming that X is m-unisolvent and Ω is a bounded, open, connected set whose boundary satisfies the cone condition.

Now set $k_G = \frac{d^{m/2}}{(2\pi)^{d/2} m!} (c_{\Omega,\kappa})^m K'_{\Omega,\kappa} \max_{|\beta|=m} \left|D^{2\beta} G(0)\right|$.

Then there exists $h_{\Omega,m} > 0$ such that for $x \in \overline{\Omega}$

$$|f(x) - (\mathcal{I}_X f)(x)| \leq k_G \sqrt{(f - \mathcal{I}_X f, f)_{w,0}} \, (h_{X,\Omega})^m \leq k_G \|f\|_{w,0} (h_{X,\Omega})^m, \tag{2.48}$$

when $h_{X,\Omega} = \sup_{\omega \in \Omega} \operatorname{dist}(\omega, X) < h_{\Omega,m}$ i.e. the order of convergence is at least m.

Further, we have the upper bound

$$|f(x) - (\mathcal{I}_X f)(x)| \leq k'_G \sqrt{(f - \mathcal{I}_X f, f)_{w,0}} \, (\operatorname{diam} \Omega)^m, \quad x \in \overline{\Omega},$$

where $k'_G = \frac{d^{m/2}}{(2\pi)^{d/2} m!} K'_{\Omega,m} \max_{|\beta|=m} \left|D^{2\beta} G(0)\right|$. The constants $c_{\Omega,m}$, $K'_{\Omega,m}$ and $h_{\Omega,m}$ only depend on Ω, m and d.

In terms of the integrals which define weight property W2 we have

$$\max_{|\beta|=m} \left|D^{2\beta} G(0)\right| \leq (2\pi)^{-d/2} \int \frac{|\xi|^{2m} d\xi}{w(\xi)}.$$

Proof. The set X is unisolvent so from Definition 66 it must have a minimal unisolvent subset, say $A = \{a_i\}_{i=1}^{M}$, which we use to define the Lagrangian operators $\mathcal{P}$ and $\mathcal{Q} = I - \mathcal{P}$ of Definition 69. Since f is a data function, $f(x) - (\mathcal{I}_X f)(x) = \mathcal{Q}(f - \mathcal{I}_X f)(x)$ and then by estimating $|\mathcal{Q}(f - \mathcal{I}_X f)(x)|$ using Lemma 71 we get

$$|f(x) - \mathcal{I}_X f(x)| \le \frac{d^{\frac{m}{2}}}{m!} \left(\sum_{i=1}^{M} |l_i(x)| \right) \left(\max_{\substack{|\beta|=m \\ z \in S_{A,x}}} \left| D^{\beta} (f - \mathcal{I}_X f)(z) \right| \right) \max_{i=1}^{M} |a_i - x|^m .$$

The next step is to consider the third factor on the last line. From part 4 of Theorem 34

$$\begin{aligned} \left| D^{\beta} (f - \mathcal{I}_X f)(z) \right| &= \left| \left(f - \mathcal{I}_X f, (-D)^{\beta} R_z \right)_{w,0} \right| \\ &\le \|f - \mathcal{I}_X f\|_{w,0} \left\| (-D)^{\beta} R_z \right\|_{w,0} \\ &\le \|f - \mathcal{I}_X f\|_{w,0} \left| \left(D^{2\beta} R_z \right)(z) \right|, \end{aligned} \tag{2.49}$$

by part 4 of Theorem 34. Finally, by part 1 of Theorem 34, $R_z = (2\pi)^{-d/2} G(\cdot - z)$ so that $\left(D^{2\beta} R_z \right)(z) = (2\pi)^{-d/2} D^{2\beta} G(0)$ and

$$\left| D^{\beta} (f - \mathcal{I}_X f)(z) \right| \le (2\pi)^{-d/2} \|f - \mathcal{I}_X f\|_{w,0} \left| D^{2\beta} G(0) \right| .$$

Thus

$$\max_{\substack{|\beta|=m \\ z \in S_{A,x}}} \left| D^{\beta} (f - \mathcal{I}_X f)(z) \right| \le (2\pi)^{-d/2} \|f - \mathcal{I}_X f\|_{w,0} \max_{|\beta|=m} \left| D^{2\beta} G(0) \right| ,$$

and so

$$|f(x) - \mathcal{I}_X f(x)| \le \frac{d^{\frac{m}{2}}}{(2\pi)^{\frac{d}{2}} m!} \|f - \mathcal{I}_X f\|_{w,0} \max_{|\beta|=m} \left| D^{2\beta} G(0) \right| \left(\sum_{i=1}^{M} |l_i(x)| \right) \max_{i=1}^{M} |a_i - x|^m .$$

To estimate the last two factors on the right side of the this equation we will need the previous Lagrangian Lemma 72. In the notation of this lemma, if $h_{X,\Omega} = \sup_{\omega \in \Omega} \operatorname{dist}(\omega, X) < h_{\Omega,m}$ then for a given x there exists a minimal unisolvent set $A = \{a_i\}_{i=1}^{M}$ such that $\operatorname{diam}(A \cup \{x\}) \le c_{\Omega,\kappa} h_{X,\Omega}$ and $\sum_{j=1}^{M} |l_j(x)| \le K'_{\Omega,m}$.

Thus $|a_i - x| \le c_{\Omega,m} h_{X,\Omega}$ and

$$\begin{aligned} |f(x) - (\mathcal{I}_X f)(x)| &\le \frac{d^{\frac{m}{2}}}{(2\pi)^{\frac{d}{2}} m!} \|f - \mathcal{I}_X f\|_{w,0} \max_{|\beta|=m} \left| D^{2\beta} G(0) \right| \; K'_{\Omega,\kappa} \left(c_{\Omega,\kappa} h_{X,\Omega} \right)^m \\ &= k_G \|f - \mathcal{I}_X f\|_{w,0} \left(h_{X,\Omega} \right)^m , \end{aligned}$$

and because $\|f - \mathcal{I}_X f\|_{w,0}^2 = (f - \mathcal{I}_X f, f)_{w,0}$ we obtain

$$|f(x) - (\mathcal{I}_X f)(x)| \le k_G \sqrt{(f - \mathcal{I}_X f, f)_{w,0}} \left(h_{X,\Omega} \right)^m , \quad x \in \Omega, \tag{2.50}$$

which is almost the required inequality. The extension of the last inequality to $x \in \overline{\Omega}$ is an easy consequence of the fact that f and $\mathcal{I}_X f$ are continuous on $\mathbb{R}^d$. The second inequality now follows directly from 2.12 and so the order of convergence of the minimal norm interpolant is at least $m = \lfloor \kappa \rfloor$.

Finally, from 1.31 we have $D^{2\beta} G(x) = (2\pi)^{-\frac{d}{2}} \int \frac{(i\xi)^{2\beta} e^{ix\xi}}{w(\xi)} d\xi$ and so $D^{2\beta} G(0) = (2\pi)^{-\frac{d}{2}} \int \frac{(i\xi)^{2\beta}}{w(\xi)} d\xi$. When $|\beta| = m$

$$(2\pi)^{\frac{d}{2}} \left| D^{2\beta} G(0) \right| \le \int \frac{\left| (i\xi)^{2\beta} \right|}{w(\xi)} d\xi \le \int \frac{|\xi|^{2|\beta|}}{w(\xi)} d\xi = \int \frac{|\xi|^{2m}}{w(\xi)} d\xi.$$

■

Thus for an arbitrary data function the order of convergence of the minimal norm interpolant is at least $\lfloor \kappa \rfloor$. Using the same technique as Corollary 60 the following *double rate* of convergence estimate can be obtained:

Corollary 75 *Under the notation and assumptions of the previous Theorem 74 the data functions $R_{x'}$ satisfy*

$$|R_{x'}(x) - (\mathcal{I}_X R_{x'})(x)| \leq (k_G)^2 (h_{X,\Omega})^{2\lfloor\kappa\rfloor}, \quad x, x' \in K, \tag{2.51}$$

when $h_{X,K} < h_G$ i.e. the order of convergence is at least $\mathbf{2}\lfloor\kappa\rfloor$.

<u>Examples: radial basis function interpolant</u>

In Subsection 1.2.3 it was shown that the weight functions of the **shifted thin-plate splines** and the **Gaussian** satisfy property W2 for all $\kappa \geq 0$ so $\lfloor\kappa\rfloor \in \mathbb{Z}_+$.

For the **Sobolev splines** the weight functions satisfy property W2 when $0 \leq \kappa < v - d/2$. Thus $\lfloor\kappa\rfloor = v - d/2 - 1$ when $v - d/2 \in \mathbb{Z}_+$ and $v - d/2 \geq 2$, and $\lfloor\kappa\rfloor = \lfloor v - d/2 \rfloor$ when $v - d/2 \notin \mathbb{Z}_+$ and $v - d/2 > 1$. Compare these with the estimates of Subsubsection 2.5.2 as summarized in Table 2.1.

<u>Examples: tensor product extended B-spline interpolant</u>

Corollary 76 *Suppose the weight function is an extended B-spline weight function with parameters $2 \leq n \leq l$. Then the order of convergence of the minimal norm interpolant to an arbitrary data function is at least $n - 1$. Further, the order of convergence to a Riesz data function R_y is at least $2n - 2$.*

Proof. Theorem 7 implies that $\lfloor\kappa\rfloor = n - 1$ and Theorem 74 implies that the order of convergence is at least $n - 1$. Corollary 75 now shows the order of convergence to a Riesz data function R_y is at least $2\lfloor\kappa\rfloor = 2n - 2$. ∎

<u>Examples: summary table</u>

Interpolant error estimates Unisolvent data: $\kappa \geq 1$		
Weight function	Parameter constraints	Convergence orders ($\lfloor\kappa\rfloor$)
Sobolev splines	$v - \frac{d}{2} = 2, 3, 4, \ldots$	$\lfloor v - d/2 \rfloor - 1$
$(v > d/2)$	$v - \frac{d}{2} > 1$, $v - \frac{d}{2} \notin \mathbb{Z}_+$	$\lfloor v - d/2 \rfloor$
Shifted thin-plate spline	-	$2, 3, 4, \ldots$
Gaussian	-	$2, 3, 4, \ldots$
Extended B-spline $(1 \leq n \leq l)$	$n \geq 2$	$n - 1$

TABLE 2.3.

2.6 Data functions and numerical results for the non-unisolvent case

In this section we will only be interested in the convergence of the interpolant to it's data function and not in the algorithm's performance as an interpolant. We will only consider the numerical experiments regarding the convergence of the interpolants generated by dilations of the one-dimensional extended B-spline basis functions 1.34 with parameters n and l. We will also restrict ourselves to one dimension so that the data density parameter can be easily calculated.

We will divide our numerical experiments into those with $n = 1$ and those with $n = 2$. For the case $n = 1$ we will consider the two cases $n = l = 1$ and $n = 1$, $l = 2$ so that derivatives are involved, and for $n = 2$ we will use $n = 2$, $l = 2$. The data region will be the interval $[-1.5, 1.5]$.

Because all the scaled extended B-spline basis weight functions have a power of $\sin x$ in the denominator we will need to derive special classes of data functions which are convenient for numerical calculations. This will include characterizing the data functions locally as Sobolev-like spaces which makes it easy to choose data functions for numerical experiments.

See also the discussion in Chapter 3 regarding the data functions generated by the *central difference* weight functions. There it is shown that locally the sets of data functions for the extended B-splines and central difference weight functions are identical.

2.6.1 Extended B-splines with $n = 1$

The case $n = 1$, $l = 1$

The hat weight function w_Λ is defined by 1.19 and was discussed in Subsections 1.2.1 and 1.2.6. This is a scaled, extended B-splines basis function with parameters $n = l = 1$ and Theorem 29 implies $\max \lfloor \kappa \rfloor = 0$.

Now suppose Π is the multivariate tensor product rectangular function defined using the 1-dimensional function $\Pi(t) = 1$ when $|t| < 1/2$ and $\Pi(t) = 0$ when $|t| > 1/2$. We will now justify using **data functions of the form** $u{*}\Pi$ where $u \in L^2(\mathbb{R}^d)$. Because $\widehat{\Pi}(t) = (2\pi)^{-1/2} \frac{\sin(t/2)}{t/2}$ and $\widehat{\Lambda}(t) = (2\pi)^{-1/2} \left(\frac{\sin(t/2)}{t/2}\right)^2$ it follows that

$$\left(\widehat{\Pi}(\xi)\right)^2 = (2\pi)^{-d/2} \widehat{\Lambda}(\xi), \quad \xi \in \mathbb{R}^d.$$

Since $\Pi \in L^1$, Young's inequality

$$\|f * g\|_r \le \|f\|_p \|g\|_q, \quad f \in L^p, g \in L^q, \ \frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}, \ 1 \le p, q, r \le \infty,$$

implies $u{*}\Pi \in L^2(\mathbb{R}^d)$ and $u{*}\Pi = (2\pi)^{-d/2} \int u(y)\,\Pi(\cdot - y)\,dy$. Further $u{*}\Pi \in L^2$ implies $u{*}\Pi \in L^1_{loc}$ and if $w_\Lambda = \frac{1}{\widehat{\Lambda}}$ is the hat weight function

$$\|u * \Pi\|_{w_\Lambda,0} = \left(\int w_\Lambda \left|\widehat{u}\,\widehat{\Pi}\right|^2\right)^{1/2} = \left(\int \frac{1}{\widehat{\Lambda}} \left|\widehat{u}\,\widehat{\Pi}\right|^2\right)^{1/2} = (2\pi)^{-\frac{d}{4}} \|u\|_2,$$

so that $u * \Pi \in X^0_{w_\Lambda}$. Now $u \in L^2$ implies $u \in L^1_{loc}$ and we can define $V \in L^1_{loc}(\mathbb{R}^d)$ by the integral

$$V(x) = \int_0^x u(t)\,dt, \quad u \in L^2(\mathbb{R}^d), \tag{2.52}$$

over the volume of the open rectangle $\mathcal{R}(0, x)$ so that

$$\begin{aligned} u * \Pi = (2\pi)^{-\frac{d}{2}} \int u(y)\,\Pi(x - y)\,dy &= (2\pi)^{-\frac{d}{2}} \int_{x-1/2}^{x+1/2} u(y)\,dy \\ &= (2\pi)^{-\frac{d}{2}} \left(V\left(x + \frac{1}{2}\right) - V\left(x - \frac{1}{2}\right)\right), \end{aligned}$$

and we note that any translate is also a data function.

Observe that by the Cauchy-Schwartz inequality $|V(x) - V(y)| = \left|\int_y^x u(t)\,dt\right| \le \|u\|_2 |x - y|^{d/2}$ so V is continuous and since $u \in L^1(\mathcal{R}(0, x))$ equation 2.52 implies that $D^{\mathbf{1}} V = u$ where $\mathbf{1} = (1, 1, \ldots, 1)$.

To obtain our 1-dimensional data function f_d we will choose

$$u = e^{-x^2},$$

for which

$$V = (2\pi)^{\frac{1}{2}} \operatorname{erf}, \qquad \|u\|_2 = 2(2\pi)^{\frac{1}{4}}, \qquad \|u * \Pi\|_{w_\Lambda,0} = 2, \tag{2.53}$$

and

$$f_d = u * \Pi = \operatorname{erf}\left(x + \frac{1}{2}\right) - \operatorname{erf}\left(x - \frac{1}{2}\right). \tag{2.54}$$

For the *double rate* convergence experiment we will use $\Lambda = \sqrt{2\pi} R_0$ as the data function. By part 2 of Theorem 6, $\kappa < 1/2$ so we must use the Type 1 error estimates of Theorems 59 and 58. These imply that if

$$G(0) - \operatorname{Re} G(x) \le C_G |x|^{2s}, \quad |x| \le h_G,$$

then there exists some $h_G > 0$ such that

$$|f_d(x) - (\mathcal{I}_X f_d)(x)| \le \|f_d\|_{w,0} \min\left\{k_G (h_{X,K})^s, \sqrt{R_0(0)}\right\}, \quad x \in K, \tag{2.55}$$

when $h_{X,K} = \max_{s \in K} \operatorname{dist}(s, X) \leq h_G$. Here $k_G = (2\pi)^{-\frac{1}{4}} \sqrt{2C_G}$, $R_0(0) = (2\pi)^{-\frac{1}{2}} G(0)$ and X is an independent data set contained in the closed, bounded, infinite data region K. From Corollary 60 we have the corresponding double order convergence estimate

$$|R_0(x) - (\mathcal{I}_X R_0)(x)| \leq \min\left\{(k_G)^2 (h_{X,K})^{2s}, R_0(0)\right\}, \quad x \in \mathbb{R}^1. \tag{2.56}$$

Since we are using the hat basis function Corollary 61 gives: $G(0) = 1$, $C_G = 1$, $s = 1/2$, $h_G = \infty$, and from 2.53 and 2.54, $\|f_d\|_{w,0} = 2$.

Numerical results

Using the functions and parameters discussed in the last subsection the four subplots displayed in Figure 2.1 each display the superposition of 20 interpolants.

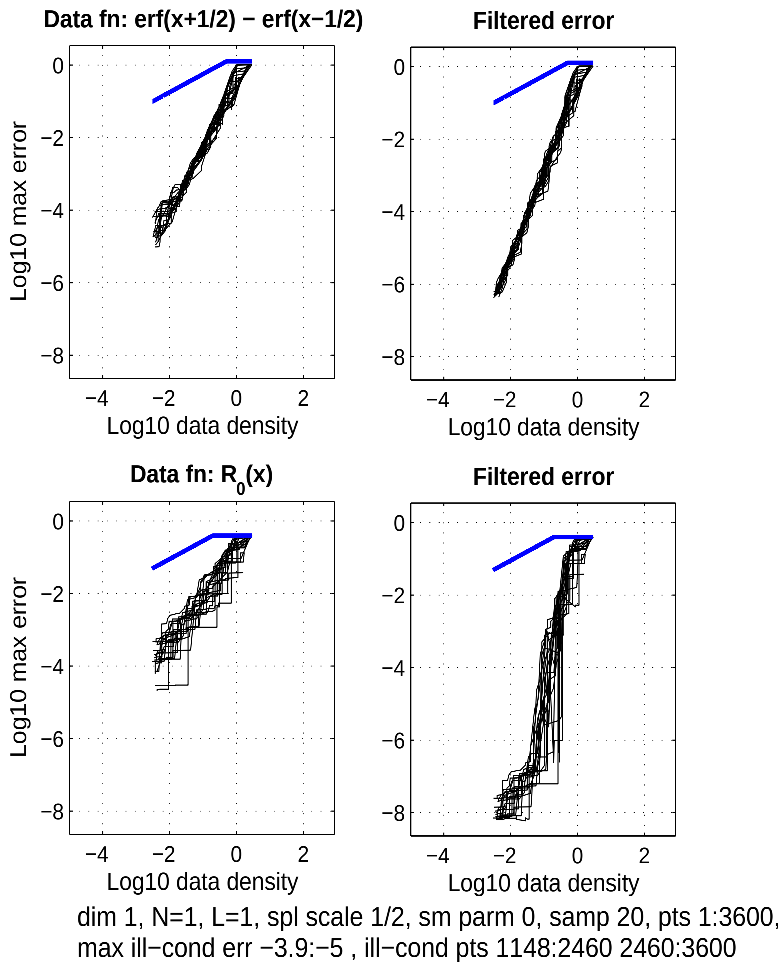

FIGURE 2.1. Interpolant convergence: extended B-spline.

The two upper subplots relate to the data function 2.54 and the lower subplots relate to the data function R_0 i.e. the Riesz representer 1.37. The right-hand subplots are filtered versions of the actual unstable interpolant. The data function is given at the top of the left-hand plots and the annotation at the bottom of the figure supplies the following additional information:

$\boxed{\text{Input parameters}}$

$\mathbf{N = L = 1}$ - the hat function is a member of the family of scaled extended B-splines with the indicated parameter values.

spl scale 1/2 - changes basis function scale (dilation): $x' = x/spl\_scale$.

sm parm 0 - the smoothing parameter is always zero for interpolation.

samp 20 - the sample size i.e. the number of test data files generated. The data function is evaluated on the interval [-1.5,1.5] using a uniform (statistical) distribution.

pts 2:3600 - specifies the smallest number of data points 2 and the largest number of data points 3600. The other values are given in exponential steps with a multiplier of approximately 1.2.

Output parameters/messages

max ill-cond err -3.5:-4.2 - this relates the ill-conditioning of the interpolation matrix to the interpolation error. A number or aster * preceding the colon refers to the data function f_d and a number or aster after the colon refers to the data function R_0. An aster * will mean there were no ill-conditioned interpolation matrices generated by the data function. Here -3.5 means that the largest (unfiltered) interpolation error for which a matrix was ill-conditioned was $10^{-4.2}$.

ill-cond pts 1234:3600 1499:2963 - the first colon-separated group corresponds to the data function f_d and indicates the smallest and largest numbers of data points for which the interpolant matrix was ill-conditioned. Asters indicate no ill-conditioning. The second colon-separated group refers to the data function R_0.

Note that all the plots shown in this document have the same format and annotations.

As mentioned above the interpolants are filtered. The filter calculates the value below which 90% of the interpolant errors lie. The filter is designed to remove 'large', isolated spikes which dominate the actual errors. The interpolation error is calculated on a grid with 300 cells applied to the domain of the data function. No filter is used for the first five interpolants because there is no instability for small numbers of data points.

As the number of points increases the numerical smoother of R_0 is observed to simplify to three very large increasingly narrow spikes at ± 1.5 and 0, and these dominate by about three orders of magnitude a residual stable error function of uniform amplitude and zero trend.

The smoother of f_d consists of intermingled spikes of various heights superimposed on a trend curve of amplitude comparable to the average spike size. The maximum spike height is at most about one order of magnitude of the average spike height. At the boundary, where the trend is near zero, there are often two spikes which are narrow w.r.t. the interior spikes.

The (blue) kinked line at the top of each subplot in Figure 2.1 is the theoretical upper bound for the error given by inequalities 2.55 or 2.56. Clearly for the data function f_d the theoretical bound of 1/2 underestimates the convergence rate by a factor of about four - assuming the filtering is valid. The situation for the data function R_0 is more complex. The double theoretical convergence rate of $2s = 1$ might suggest that the actual interpolant converges twice as quickly for the Riesz data function and the filtering reinforces this suspicion. However, there is no nice linear decrease observed and the error of each interpolant decreases in large steps to the stable state.

It is also very interesting to note that the theoretical upper error bound for the data function R_0 is able to take into account quite closely the instability of the interpolant.

The case $n = 1$, $l = 2$:

Amongst other things, the following result will equip us with some tools to generate data functions for the extended B-splines data spaces, data functions for which the X_w^0 norm can be calculated This result is closely related to the calculations done above for the hat function.

In the sequel translations will be defined by $\tau_c f(x) = f(x - c)$, $c \in \mathbb{R}^d$, and dilations given by $\sigma_\lambda f(x) = f(x/\lambda)$ where $\lambda \in \mathbb{R}^1$ and $\lambda > 0$. If $x, y \in \mathbb{R}^d$ then $x.y = (x_i y_i)$ denotes the *component-wise product* of x and y whereas the scalar or inner product is denoted by xy or (x, y).

Theorem 77 ***Central difference operators*** *and* ***data functions for B-splines***

Suppose for $c \in \mathbb{R}^d$ we define the central difference operator δ_c^α by $\delta_c^\alpha = \delta_{c_1}^{\alpha_1 \mathbf{e}_1} \delta_{c_2}^{\alpha_2 \mathbf{e}_2} \ldots \delta_{c_d}^{\alpha_d \mathbf{e}_d}$ where

$$\delta_{c_k}^{\mathbf{e}_k} = \tau_{-\frac{c_k}{2}\mathbf{e}_k} - \tau_{\frac{c_k}{2}\mathbf{e}_k}, \quad k = 1, \ldots, d. \tag{2.57}$$

and use the abbreviation $\delta_b^\alpha = \delta_{b\mathbf{1}}^\alpha$ when $b \in \mathbb{R}^1$. Then it follows that

$$\delta_c^\alpha = \sum_{\beta \leq \alpha} (-1)^{|\beta|} \binom{\alpha}{\beta} \tau_{(2\beta - \alpha).\frac{c}{2}}. \tag{2.58}$$

Next, suppose w is the extended B-spline weight function with parameters n and l given by 1.20, and that in the sense of distributions $D^\alpha U \in L^2\left(\mathbb{R}^d\right)$ when $\alpha \leq n$.

Then, regarding data functions, if $f_d = \delta_2^{l\mathbf{1}} U$ then $f_d \in X_w^0$ and

$$\|f_d\|_{w,0} = 2^{ld} \left\|D^{n\mathbf{1}} U\right\|_2. \tag{2.59}$$

Further:

1. *Translations are isometric isomorphisms from X_w^0 to X_w^0.*
2. *If $m > 0$ is an integer then the dilation σ_m is a continuous mapping from X_w^0 to X_w^0.*
3. *The embedding $X_w^0 \hookrightarrow C_B^{(n-1)}$ is continuous when $C_B^{(n-1)}$ is endowed with the supremum norm $\max\limits_{|\alpha| \leq n-1} \|D^\alpha u\|_\infty$.*
4. *Suppose $v \in C_B^\infty\left(\mathbb{R}^d\right)$. If v has periods $\mathbf{1}$ then $v\delta_2^{l\mathbf{1}} U = \delta_2^{l\mathbf{1}}(vU) \in X_w^0$ for all l. If v has periods $2\mathbf{1}$ then $v\delta_2^{l\mathbf{1}} U = \delta_2^{l\mathbf{1}}(vU) \in X_w^0$ when l is even.*

Proof. Equation 2.58 can be easily proved in one dimension and then for arbitrary dimension using the definitions of $\binom{\alpha}{\beta}$ and $\beta \leq \alpha$. Since $\widehat{\delta_2^{\mathbf{e}_k} U} = e^{i\xi_k}\widehat{U} - e^{-i\xi_k}\widehat{U} = (2i \sin \xi_k)\widehat{U}$ it follows that $\widehat{f_d} = (2i \sin \xi_k)^{l\mathbf{1}} \widehat{U}$. Now using the Cauchy-Schwartz theorem it is easy to show that any L^2 function is L^1_{loc}. Further, Plancherel's theorem states that $\left\|\widehat{U}\right\|_2 = \|U\|_2$ and so $\widehat{U} \in L^1_{loc}$ and $\widehat{f_d} \in L^1_{loc}$. Again by Plancherel's theorem $D^\beta U \in L^2$ implies $\left\|D^\beta U\right\|_2^2 = \left\|\widehat{D^\beta U}\right\|_2^2 = \int \xi^{2\beta} \left|\widehat{U}\right|^2$. Consequently

$$\begin{aligned}
\|f_d\|_{w,0}^2 = \int w \left|\widehat{f_d}\right|^2 &= \int \frac{\xi^{2n\mathbf{1}}}{(\sin \xi_k)^{2l\mathbf{1}}} \left|(2i \sin \xi_k)^{l\mathbf{1}} \widehat{U}\right|^2 \\
&= 2^{2ld} \int \xi^{2n\mathbf{1}} \left|\widehat{U}\right|^2 = 2^{2ld} \left\|D^{n\mathbf{1}} U\right\|_2^2,
\end{aligned} \tag{2.60}$$

and $f_d \in X_w^0$ which establishes 2.59.

Part 1 The stated properties of translations τ_c follow directly from the facts that $\tau_c \tau_{-c} = \tau_{-c}\tau_c = 1$ and $|\widehat{\tau_c u}| = |\widehat{u}|$.

Part 2 Regarding the dilations, by using two changes of variable we obtain

$$\begin{aligned}
\|\sigma_m f_d\|_{w,0}^2 = \int w(t) \left|\widehat{\sigma_m f_d}\right|^2 (t)\, dt &= m^{2d} \int w(t) \left|\widehat{f_d}(mt)\right|^2 dt \\
&= m^d \int w\left(\frac{t}{m}\right) \left|\widehat{f_d}(t)\right|^2 dt \\
&= m^d \int \frac{\left(\frac{t}{m}\right)^{2n\mathbf{1}}}{\left(\sin \frac{t_k}{m}\right)^{2l\mathbf{1}}} \left|\widehat{f_d}(t)\right|^2 dt \\
&= \frac{1}{m^{(2n-1)d}} \int \frac{t^{2n\mathbf{1}}}{\left(\sin \frac{t_k}{m}\right)^{2l\mathbf{1}}} \left|\widehat{f_d}(t)\right|^2 dt \\
&= \frac{1}{m^{(2n-1)d}} \int \frac{(\sin t_k)^{2l\mathbf{1}}}{\left(\sin \frac{t_k}{m}\right)^{2l\mathbf{1}}} \frac{t^{2n\mathbf{1}}}{(\sin t_k)^{2l\mathbf{1}}} \left|\widehat{f_d}(t)\right|^2 dt \\
&= \frac{1}{m^{(2n-1)d}} \int \left(\frac{\sin t_k}{\sin \frac{t_k}{m}}\right)^{2l\mathbf{1}} w(t) \left|\widehat{f_d}(t)\right|^2 dt \\
&\leq m^{(2(l-n)+1)d} \|f_d\|_{w,0}^2,
\end{aligned} \tag{2.61}$$

since $\left|\frac{\sin s_k}{\sin \frac{s_k}{m}}\right| \leq m$ for all $s_k \in \mathbb{R}^1$ and all integers $m \geq 1$. This proves continuity.

Part 3 From part 6 Theorem 34 the embedding $X_w^0 \hookrightarrow C_B^{(\kappa)}$ is continuous and 1.21 allows us to choose $\kappa = n - 1$.

Part 4 Firstly, $U \in L^2$ implies $vU \in L^2$. Further, since $L^2 \subset S'$ we have by virtue of Leibniz' formula for tempered distributions S'

$$D^{\alpha}(vU) = \sum_{\beta \leq \alpha} \binom{\alpha}{\beta} D^{\alpha-\beta} v D^{\beta} U, \quad U \in L^2,\ v \in C_B^{\infty}.$$

Now $D^{\alpha}U \in L^2$ for $\alpha \leq n$, so that $D^{\alpha}(vU) \in L^2$ when $\alpha \leq n$. Since $f_d \in X_w^0 \subset S' \subset \mathcal{D}'$, if v has periods $\mathbf{1}$ then in the sense of distributions, in one dimension 2.58 implies

$$v\delta_2^{l\mathbf{e}_k} U = v \sum_{j=0}^{l} (-1)^j \binom{l}{j} U(\cdot + (l-2j)\, e_k) = \sum_{j=0}^{l} (-1)^j \binom{l}{j} (vU)(\cdot + (l-2j)\, e_k) = \delta_2^{l\mathbf{e}_k}(vU).$$

Consequently $vf_d = v\delta_2^{l\mathbf{1}} U = \delta_2^{l\mathbf{1}}(vU)$ and thus 2.59 of this theorem implies $vf_d \in X_w^0$. Finally, it is now clear that if v has periods $2\mathbf{1}$ and l is even then $vf_d = \delta_2^{l\mathbf{1}}(vU)$. ■

Remark 78

1. ***Norms of data functions*** *Further to the comment preceding the last theorem, if $D^{\alpha}U \in L^2(\mathbb{R}^d)$ for $\alpha \leq n$ then $f_d = \delta_2^{l\mathbf{1}} U$ is a data function with norm $2^{ld} \left\| D^{n\mathbf{1}} U \right\|_2$ and translating this function does not change the norm. In addition, if $v \in C_B^{\infty}(\mathbb{R}^d)$ has periods $\mathbf{1}$ when l is odd or ν has periods $2\mathbf{1}$ when l is even, then $vf_d = \delta_2^{l\mathbf{1}}(vU)$ is also a data function with norm $2^{ld} \left\| D^{n\mathbf{1}}(vU) \right\|_2$.*

2. ***Regarding part 4 of the last theorem*** *The condition $v \in C_B^{\infty}(\mathbb{R}^d)$ can be weakened to $v \in H^{n,\infty}(\mathbb{R}^d)$ where*
   $H^{n,\infty}(\mathbb{R}^d) = \{f \in L^{\infty} : D^{\alpha} f \in L^{\infty},\ 0 \leq |\alpha| \leq n\}$ is the L^{∞} Sobolev space introduced below in Definition 80. This can be proved using the Leibniz formula $D(vf) = vDf + (Dv)\, f, \quad v \in H^{n,\infty}$, $f \in H^n$ which can in turn be demonstrated by adapting the proof of Theorem 14.2, Section 14.14 Petersen [11].

Our basis function is the (unscaled) 1-dimensional extended B-spline G_1 with parameters $n = 1$ and $l = 2$ given by 1.34 i.e.

$$G_1(t) = (-1)^{l-n} \tfrac{(2\pi)^{l/2}}{2^{2(l-n)+1}} \left(D^{2(l-n)} \left((*\Lambda)^l \right) \right) \left(\tfrac{t}{2}\right) = -\tfrac{\pi}{4} \left(D^2 (\Lambda * \Lambda) \right) \left(\tfrac{t}{2}\right)$$

for $t \in \mathbb{R}^1$. But

$$\begin{aligned} D^2(\Lambda * \Lambda) = \Lambda * D^2 \Lambda &= \Lambda * (\delta(\cdot + 1) - 2\delta + \delta(\cdot - 1)) \\ &= \frac{1}{\sqrt{2\pi}} (\Lambda(\cdot + 1) - 2\Lambda + \Lambda(\cdot - 1)), \end{aligned} \tag{2.62}$$

so that

$$G_1(t) = -\frac{\sqrt{2\pi}}{8} \left(\Lambda\left(\tfrac{t}{2} + 1\right) - 2\Lambda\left(\tfrac{t}{2}\right) + \Lambda\left(\tfrac{t}{2} - 1\right) \right),$$

and hence

$$DG_1(t) = -\frac{\sqrt{2\pi}}{16} \left(\Lambda'\left(\tfrac{t}{2} + 1\right) - 2\Lambda'\left(\tfrac{t}{2}\right) + \Lambda'\left(\tfrac{t}{2} - 1\right) \right),$$

i.e. $\|DG_1\|_{\infty} = \frac{3}{16}\sqrt{2\pi}$. Now by Theorem 43

$$G_1(0) - G_1(t) \leq \|DG_1\|_{\infty} |t|, \quad x \in \mathbb{R}^1,$$

which means that

$$G_1(0) = \frac{\sqrt{2\pi}}{4}, \quad C_G = \|DG_1\|_{\infty} = \frac{3}{16}\sqrt{2\pi}, \quad s = \frac{1}{2}, \quad h_G = \infty.$$

With reference to the last theorem we will choose the bell-shaped data function

$$f_d = \delta_2^2 U \in X_w^0, \tag{2.63}$$

where

$$U(x)=\frac{e^{-k_{1,2}x^2}}{\delta_2^2\left(e^{-k_{1,2}x^2}\right)(0)}=\frac{e^{-k_{1,2}x^2}}{2\left(1-e^{-4k_{1,2}}\right)},\quad k_{1,2}=0.3, \tag{2.64}$$

so that

$$\|f_d\|_{w,0}=4\|DU\|_2=\sqrt[4]{2\pi}\frac{\sqrt[4]{4k_{1,2}}}{1-e^{-4k_{1,2}}}.$$

The interpolation error estimates given by 2.55 and 2.56 are now

$$|f_d(x)-(\mathcal{I}_X f_d)(x)|\leq\|f_d\|_{w,0}\min\left\{k_G(h_{X,K})^s,\sqrt{R_0(0)}\right\},\quad x\in K,$$

and

$$|R_0(x)-(\mathcal{I}_X R_0)(x)|\leq\min\left\{(k_G)^2(h_{X,K})^{2s},R_0(0)\right\},\quad x\in\mathbb{R}^1,$$

respectively, where $k_G=(2\pi)^{-\frac{1}{4}}\sqrt{2C_G}$ and $R_0(0)=(2\pi)^{-\frac{1}{2}}G_1(0)$.

For the theory developed in the previous sections it was convenient to use the simple, unscaled weight function definition 1.20. However, I have observed that scaling can significantly improve the performance of the basis function interpolant and so I will present the following theorem for the scaled, extended B-splines.

Theorem 79 *Suppose* $\widetilde{G}(x)=\prod_{k=1}^{d}\widetilde{G}_1(x_k)$ *where* $\widetilde{G}_1=(-1)^{l-n}D^{2(l-n)}\left((*\Lambda)^l\right)$ *and* Λ *is the univariate hat function and* n,l *are integers such that* $1\leq n\leq l$.

For given $\lambda>0$, $\widetilde{G}(\lambda x)$ *is called a scaled extended B-spline basis function. The corresponding weight function is* $\widetilde{w}_\lambda(t)=(2\lambda a)^d w\left(\frac{t}{2\lambda}\right)$ *where* $a=\frac{(2\pi)^{l/2}}{2^{2(l-n)+1}}$ *and* w *is the extended B-spline weight function 1.20 with parameters* n,l. *Indeed,* $\widetilde{w}_\lambda$ *has property W2 for* κ *iff* w *has property W2 for* κ. *Further*

$$\widetilde{G}(0)-\widetilde{G}(\lambda x)\leq 2\lambda a^{-d}\sqrt{d}G_1(0)^{d-1}\|DG_1\|_\infty|x|,\quad x\in\mathbb{R}^d. \tag{2.65}$$

Finally, if $f_d\in X_w^0$ *and* $g_d(x)=f_d(2\lambda x)$, *it follows that* $g_d\in X_{\widetilde{w}_\lambda}^0$ *and* $\|g_d\|_{\widetilde{w}_\lambda,0}=a^{d/2}\|f_d\|_{w,0}$.

Proof. From 1.34, $\widetilde{G}(x)=a^{-d}G(2x)$, where G is the extended B-spline basis function with parameters n and l. Hence by Theorem 23 the corresponding weight function is $w_\lambda(t)=(2\lambda a)^d w\left(\frac{t}{2\lambda}\right)$ with property W2 for κ. By Theorem 43

$$G(0)-G(x)\leq\sqrt{d}G_1(0)^{d-1}\|DG_1\|_\infty|x|,\quad x\in\mathbb{R}^1,$$

and because $\widetilde{G}(\lambda x)=a^{-d}G(2\lambda x)$ we have

$$\begin{aligned}\widetilde{G}(0)-\widetilde{G}(\lambda x)=a^{-d}(G(0)-G(2\lambda x))&\leq a^{-d}\sqrt{d}G_1(0)^{d-1}\|DG_1\|_\infty|2\lambda x|\\&=2\lambda a^{-d}\sqrt{d}G_1(0)^{d-1}\|DG_1\|_\infty|x|.\end{aligned}$$

Finally, $\widehat{g_d}(t)=(2\lambda)^{-d}\widehat{f_d}\left(\frac{t}{2\lambda}\right)$ and so

$$\|g_d\|_{w_\lambda,0}^2=\int(2\lambda a)^d w\left(\frac{t}{2\lambda}\right)\left|(2\lambda)^{-d}\widehat{f_d}\left(\frac{t}{2\lambda}\right)\right|^2dt=a^d\int w(s)\left|\widehat{f_d}(s)\right|^2ds=a^d\|f_d\|_{w,0}^2.$$

■

Numerical results

Using the functions and parameters discussed in the last subsection the four subplots displayed in Figure 2.2 each display the superposition of 20 interpolants. The two upper subplots relate to the data function 2.63 and the lower subplot relates to the data function R_0 i.e. the Riesz representer 1.37. The right-hand subplots are filtered versions of the actual unstable interpolant. The four subplots of Figure 2.2 each display the superposition of 20 interpolants.

As the number of points increases the numerical interpolant of f_d gradually simplifies to two spikes at the end points of the data interval with stable errors in between. The interpolant of R_0 is observed

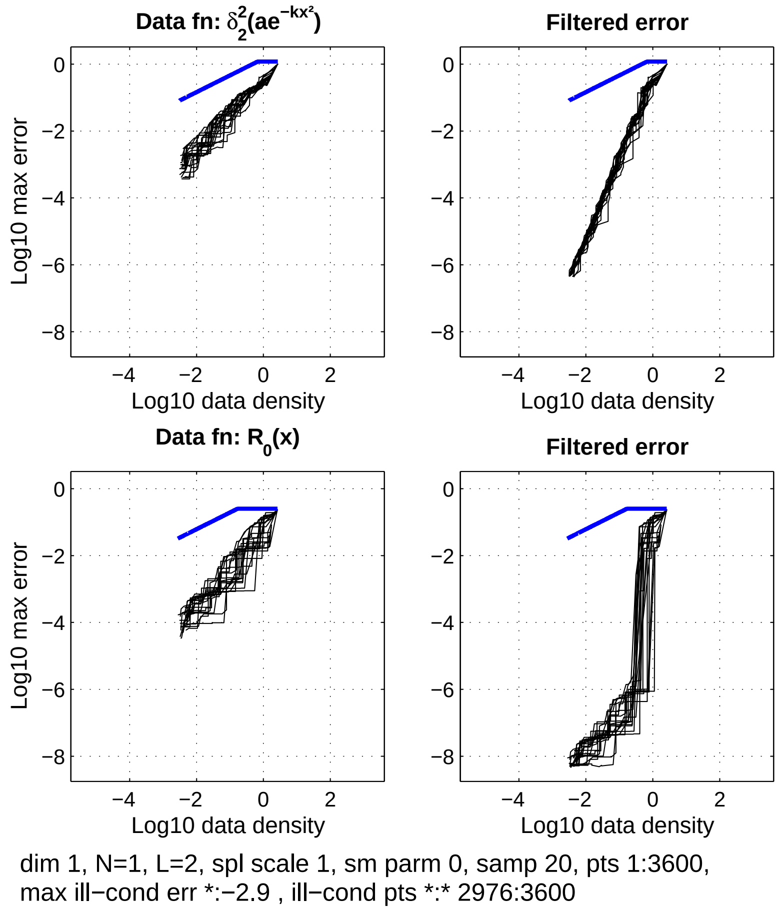

FIGURE 2.2. Interpolant convergence: extended B-spline.

to simplify to a large increasingly narrow spike at zero and this dominates a stable error function with absolute value of the order of 10^{-7}.

The (blue) kinked line above each interpolant at the top of each subplot in Figure 2.2 is the estimated upper bound for the error given by the inequalities 2.55 or 2.56. Clearly for the data function f_d the theoretical bound of $1/2$ underestimates the convergence rate by a factor of approximately four - assuming the filtering is valid. The situation for the data function R_0 is more complex. The double theoretical convergence rate of $2s = 1$ might suggest that the actual interpolant converges twice as quickly for the Riesz data function and the filtering reinforces this suspicion. However, there is no nice linear decrease observed and the error of each interpolant decreases in a large step and then smaller steps to the stable state.

It is also very interesting to note that the theoretical upper error bound for the data function R_0 is able to take into account quite closely the instability of the interpolant.

2.6.2 Extended B-splines with $n = 2$

Since $n \geq 2$, Table 2.2 tells us we can use the Type 2 error estimates of Theorem 65 and the estimate of Theorem 58.

The case: $n = 2$, $l = 2$

From Corollary 79 the function

$$G_{2,2}(x) = \frac{(\Lambda * \Lambda)(2x)}{(\Lambda * \Lambda)(0)} = \frac{3\sqrt{2\pi}}{2}(\Lambda * \Lambda)(2x),$$

is a scaled extended B-spline basis function with parameters $n = 2$, $l = 2$ such that $\operatorname{supp} G_{2,2} = [-1, 1]$ and $G_{2,2}(0) = 1$. From 1.34 the scaling factor is $\lambda = 4$. To calculate $G_{2,2}$ we use the convenient formula

$$G_{2,2}(x) = (1+x)^2 \Lambda(2x+1) + \left(1 - 2x^2\right) \Lambda(2x) + (1-x)^2 \Lambda(2x-1), \tag{2.66}$$

derived by observing that we require the symmetric equivalents of the B-splines b^k studied, for example, in Chapter 3 of Höllig [5]: Noting that $\Lambda(x) = b^1(x+1)$, the symmetric equivalents of the convolution formula of box 3.11 and of the formulas of boxes 3.3, 3.4 can be used to derive 2.66. With reference to Theorem 77 choose the data function 2.63, 2.64 and for double rate convergence experiments we will use $R_0 = (2\pi)^{-\frac{1}{2}} G_{2,2}$ as the data function.

Since $n \geq 2$, Table 2.2 tells us we can use the Type 2 error estimates of Theorem 65 and the estimate of Theorem 58. These imply that

$$|f_d(x) - (\mathcal{I}_X f_d)(x)| \leq \|f_d\|_{w,0} \min\left\{k_G h_{X,K}, \sqrt{R_0(0)}\right\}, \quad x \in K,\ f_d \in X_w^0, \tag{2.67}$$

and

$$|R_0(x) - (\mathcal{I}_X R_0)(x)| \leq \min\left\{(k_G)^2 (h_{X,K})^2, R_0(0)\right\}, \quad x \in \mathbb{R}^1, \tag{2.68}$$

where $h_{X,K} = \max\limits_{s \in K} \operatorname{dist}(s, X)$ and in addition

$$\begin{aligned} k_G &= (2\pi)^{-\frac{1}{4}} \sqrt{-G_{2,2}(0)^{d-1} D^2 G_{2,2}(0)} \sqrt{d} = (2\pi)^{-\frac{1}{4}} \sqrt{-D^2 G_{2,2}(0)}, \\ R_0(0) &= (2\pi)^{-\frac{1}{2}} G_{2,2}(0) = (2\pi)^{-\frac{1}{2}}, \end{aligned}$$

and $K = [-1.5, 1.5]$.

Using 2.62 we have $-D^2 G_{2,2}(0) = -6\sqrt{2\pi} D^2 (\Lambda * \Lambda)(0) = 6\sqrt{2\pi} \frac{2}{\sqrt{2\pi}} = 12$ so that $k_G = (2\pi)^{-\frac{1}{4}} \sqrt{12}$.

It remains to calculate $\|f_d\|_{w,0}$. But from 2.63 and Plancherel's theorem

$$\|f_d\|_{w,0} = 2^{ld} \left\|D^{n\mathbf{1}} U\right\|_2 = 4 \left\|D^2 U\right\|_2 = 4 \left\|\xi^2 \widehat{U}\right\|_2.$$

By 2.64, $U(x) = \frac{e^{-k_{1,2} x^2}}{2\left(1 - e^{-4k_{1,2}}\right)}$ so $\widehat{U}(\xi) = \frac{1}{2\left(1 - e^{-4k_{1,2}}\right)} \frac{1}{\sqrt{k_{1,2}}} e^{-\frac{\xi^2}{k_{1,2}}}$ and

$$\begin{aligned} \left\|\xi^2 \widehat{U}\right\|_2 &= \frac{4}{2\left(1 - e^{-4k_{1,2}}\right)\sqrt{k_{1,2}}} \left(\int \xi^4 e^{-\frac{2\xi^2}{k_{1,2}}} d\xi\right)^{1/2} \\ &= \frac{2}{\left(1 - e^{-4k_{1,2}}\right)\sqrt{k_{1,2}}} \frac{k_{1,2}}{2} \sqrt[4]{\frac{k_{1,2}}{2}} \left(\int \xi^4 e^{-\xi^2} d\xi\right)^{1/2} \\ &= \frac{2}{\left(1 - e^{-4k_{1,2}}\right)\sqrt{k_{1,2}}} \frac{k_{1,2}}{2} \sqrt[4]{\frac{k_{1,2}}{2}} \left(\frac{3}{4}\sqrt{\pi}\right)^{1/2} \\ &= \frac{1}{2\left(1 - e^{-4k_{1,2}}\right)\sqrt{k_{1,2}}} k_{1,2} \sqrt[4]{k_{1,2}} \left(\frac{9\pi}{2}\right)^{1/4} \\ &= \frac{1}{2} \sqrt[4]{\frac{9\pi}{2}} \frac{(k_{1,2})^{3/4}}{1 - e^{-4k_{1,2}}}, \end{aligned}$$

and hence $\|f_d\|_{w,0} = 2\sqrt[4]{\frac{9\pi}{2}} \frac{(k_{1,2})^{3/4}}{1 - e^{-4k_{1,2}}} = \sqrt[4]{72\pi} \frac{(k_{1,2})^{3/4}}{1 - e^{-4k_{1,2}}}$. To summarize:

$$k_G = \frac{\sqrt{12}}{(2\pi)^{1/4}}; \quad \|f_d\|_{w,0} = \sqrt[4]{72\pi} \frac{(k_{1,2})^{3/4}}{1 - e^{-4k_{1,2}}}, \quad k_{1,2} = 0.3; \quad h_G = \infty.$$

Numerical results

Using the functions and parameters derived above the 4 subplots of Figure 2.3 were generated. Each plot displays the superposition of 20 interpolants with unfiltered output on the left and filtered interpolants on

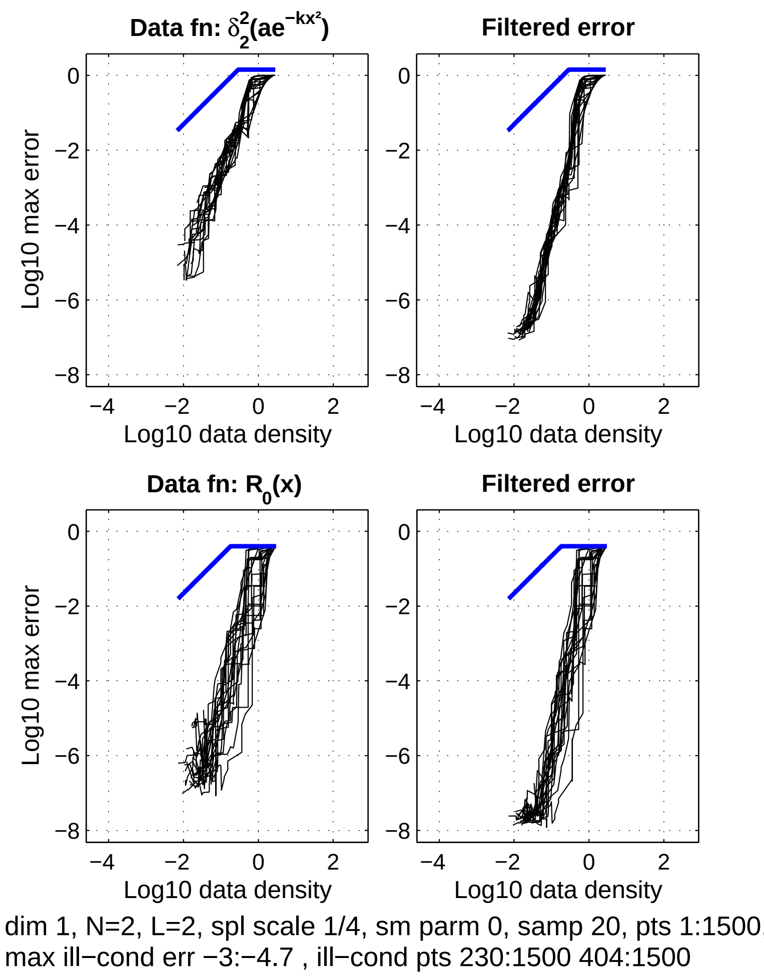

FIGURE 2.3. Interpolant convergence: extended B-spline.

the right. Above the interpolants is a (blue) kinked line whose slope indicates the theoretically predicted rate of convergence given by inequalities 2.67 and 2.68. There is also an adjacent line which represents an estimate of the slope implied by the filtered interpolant. For $f_d = \delta_2^2 \frac{e^{-k_{1,2}x^2}}{2\left(1-e^{-4k_{1,2}}\right)}$ the estimated rate of convergence is $n-1=1$ and for R_0 the estimated convergence rate is $2n-2=2$.

As the number of points increases both numerical interpolants were observed to consist of spikes of various sizes but no dominant spikes.

The (blue) kinked line above each interpolant at the top of each subplot in Figure 2.3 is the theoretical upper bound for the error given by inequalities 2.67 or 2.68. Clearly for the data function f_d the theoretical bound of 1 underestimates the convergence rate by a factor of about 3 - assuming the filtering is valid. For for the data function R_0 a theoretical bound of 2 underestimates a convergence rate of about 5. The double theoretical convergence rate of **2** might suggest that the actual interpolant converges twice as quickly for the Riesz data function and the filtering reinforces this suspicion.

2.7 The restriction spaces $X_w^0(\Omega)$

Suppose $\Omega \subset \mathbb{R}^d$ is bounded and w is a scaled, extended B-spline weight function. In this section we will use the central difference operators 2.58 to show that the local restriction spaces $X_w^0(\Omega)$ contain the same set of functions as a special, local, Sobolev-like space, denoted $H_s^m(\Omega)$. We will then extend the class of weight functions to which these characterization results are applicable. This is useful numerically

because it is easy to generate functions in $H_s^m(\Omega)$. However, calculating the norm of an extension to X_w^0 may be difficult.

We start with some definitions and in these definitions we use the notation $\alpha \le m$ where m is a non-negative integer and $\alpha \in \mathbb{R}^d$ is a multi-index. This notation means:

$$\{\alpha : \alpha \le m\} = \{\alpha : \alpha_k \le m \ for \ k = 1, \dots, m\}.$$

See Definition 221 of the Appendix.

Our next important data function result Theorem 90 will require some L^2 Sobolev space theory based on the comprehensive study [1] by Adams.

Definition 80 ***Sobolev spaces*** $H^{m,\infty}(\Omega)$, $H^{m,\infty}$, $H^m(\Omega)$, H^m, $m = 0, 1, \dots$

For any open set $\Omega \subset \mathbb{R}^d$ and in the sense of distributions:

$$H^{m,\infty}(\Omega) = \{u \in L^\infty(\Omega) : D^\alpha u \in L^\infty(\Omega) \ \ for \ \ |\alpha| \le m\},$$
$$H^{m,\infty} = H^{m,\infty}\left(\mathbb{R}^d\right),$$

with respective norms $\|u\|_{m,\infty,\Omega} = \max\limits_{|\alpha|\le m} \sup\limits_{x\in\Omega} |D^\alpha u(x)|$ *and* $\|u\|_{m,\infty} = \|u\|_{m,\infty,\mathbb{R}^d}$.

$$H^m(\Omega) = \left\{u \in L^2(\Omega) : D^\alpha u \in L^2(\Omega) \ \ for \ \ |\alpha| \le m\right\},$$
$$H^m = H^m\left(\mathbb{R}^d\right),$$

with norms $\|u\|_{m,2,\Omega} = \sum\limits_{|\alpha|\le m} \left(\int_\Omega |D^\alpha u|^2\right)^{1/2}$ *and* $\|u\|_{m,2} = \|u\|_{m,2,\mathbb{R}^d}$.

We will also need the following Sobolev-like Hilbert spaces:

Definition 81 ***The spaces*** $H_s^m(\Omega)$, H_s^m, $H_{s0}^m(\Omega)$, H_{s0}^m, $m = 0, 1, 2, \dots$

$$H_s^m(\Omega) = \left\{u \in L^2(\Omega) : D^\alpha u \in L^2(\Omega) \ \ for \ \alpha \le m\right\},$$
$$H_s^m = H_s^m\left(\mathbb{R}^d\right),$$
$$H_{s0}^m(\Omega) = \{u \in H_s^m(\Omega) : \operatorname{supp} u \ is \ bounded\},$$
$$H_{s0}^m = H_{s0}^m\left(\mathbb{R}^d\right).$$

with norms $|||u|||_{m,2,\Omega} = \sum\limits_{\alpha\le m} \left(\int_\Omega |D^\alpha u|^2\right)^{1/2}$ *and* $|||u|||_{m,2} = |||u|||_{m,2,\mathbb{R}^d}$.

Remark 82 *The work by Adams [1] is devoted to Sobolev space theory and a review of this study shows that many of it's results can be extended to the spaces H_s^m, which are clearly Sobolev spaces in the case of one dimension. In particular:*

1. ***Theorem 3.2*** *proves $H_s^m(\Omega)$ is a Banach space.*
2. ***Theorem 3.5*** *shows that $H_s^m(\Omega)$ is a separable, reflexive, uniformly convex Hilbert space with inner products defined in terms of their norms.*
3. ***Lemma 3.15*** *holds.*
4. ***Theorem 3.18*** *proves using Lemma 3.15 that if Ω has the segment property (§4.2) then the space of restrictions $\{\phi|_\Omega : \phi \in C_0^\infty\}$ is dense in $H_s^m(\Omega)$.*
5. ***Theorem 4.32 (The Calderón extension operator)*** *If Ω has a strengthened uniform cone property (§4.4) then using Theorem 3.18 it can be shown that there is a continuous linear extension $\mathcal{E}_\Omega : H_s^m(\Omega) \to H_s^m$ - termed an* ***m-simple extension*** *- which has a continuity constant dependent only on m and d. Note that when the boundary is bounded the strengthened cone property reduces to the uniform cone property.*
6. ***Theorem 4.26*** *holds. Using a reflection technique a* ***strong m-extension*** *operator is constructed (k-simple for $k \le m$) from $H_s^m(\Omega)$ to H_s^m, where Ω is a half-space in $\mathbb{R}^d$ or an open set having the uniform C^m-regularity property (§4.6) and bounded boundary.*

7. ***Theorem 4.28*** *holds. Using a reflection technique a* ***total extension*** *operator is constructed i.e. the operator is a strong m-extension operator for all* $m \geq 1$.

8. *The above extension operators commute with translations and dilations in the sense that* $\tau_c \mathcal{E}_\Omega = \mathcal{E}_{\tau_c(\Omega)} \tau_c$ *and* $\sigma_\lambda \mathcal{E}_\Omega = \mathcal{E}_{\sigma_\lambda(\Omega)} \sigma_\lambda$.

I next present further results concerning the $H_s^m(\Omega)$ spaces. Observe that the proof of part 4 uses both L^1 and L^2 Fourier transform theory and introduces an equivalent Fourier transform norm for H_s^m.

Lemma 83 *The space* $H_s^m(\Omega)$ *also has the following properties:*

1. *The* ***translation*** *operator* τ_c *is an isometric isomorphism from* $H_s^m(\Omega)$ *to* $H_s^m(\Omega - c)$.

   The ***dilation*** *operator* σ_λ *is a homeomorphism from* $H_s^m(\Omega)$ *to* $H_s^m(\Omega/\lambda)$.

2. *If we define* $C_0^\infty(\Omega) = \{\phi \in C_0^\infty : \operatorname{supp} \phi \subset \Omega\}$ *then* ***multiplication by a*** $C_0^\infty(\Omega)$ ***function*** *is a continuous mapping from* $H_s^m(\Omega)$ *to* $H_s^m(\Omega)$. *In fact there exists a constant* $c_{m,d}$*, depending only on* m *and the dimension* d*, such that*

$$|||\phi u|||_{m,2,\Omega} \leq c_{m,d} \|\phi\|_{m,\infty} |||u|||_{m,2,\Omega}, \quad \phi \in C_0^\infty(\Omega), \ u \in H_s^m(\Omega). \tag{2.69}$$

3. ***Extension*** *Suppose* Ω *permits an extension operator* $\mathcal{E}_\Omega$ *described in parts 5 to 7 of Remark 82. Then given* $\varepsilon > 0$ *there exists a continuous linear extension mapping* $E_{\Omega;\varepsilon} : H_s^m(\Omega) \to H_s^m(\Omega_\varepsilon)$*, and which has a continuity constant depending only on* ε*,* m *and* d*.*

   This extension commutes with translations and dilations in the sense that $\tau_c E_{\Omega;\varepsilon} = E_{\tau_c(\Omega);\varepsilon} \tau_c$ *and* $\sigma_\lambda E_{\Omega;\varepsilon} = E_{\sigma_\lambda(\Omega);\varepsilon/\lambda} \sigma_\lambda$.

4. ***Embedding*** *Suppose* Ω *has the properties specified in part 3. Then* $H_s^m \subset C_B^{(m-1)}$ *and* $H_s^m(\Omega) \subset C_B^{(m-1)}(\Omega)$ *and these inclusions are continuous under the corresponding* $H^{m-1,\infty}$ *supremum norms.*

Proof. Part 1 $H_s^m(\Omega)$ is closed under translations and dilations, the inverses of translations and dilations are translations and dilations respectively, and the continuity calculations are simple changes of variable.

Part 2 Use the Leibniz theorem for distributions to expand $D^\alpha(\phi u)$ for each $1 \leq \alpha \leq m$ to obtain: $\|D^\alpha(\phi u)\|_2 \leq c'_{m,d} \|\phi\|_{m,\infty} |||u|||_{m,2,\Omega}$, where $c'_{m,d}$ depends only on m and d.

Part 3 Using Lemma 17 we construct a function $\phi_{\Omega;\varepsilon} \in C_0^\infty(\Omega_\varepsilon)$ such that $0 \leq \phi_{\Omega;\varepsilon} \leq 1$, $\phi_{\Omega;\varepsilon} = 1$ on $\Omega_{\varepsilon/4}$ and $\operatorname{supp} \phi_{\Omega;\varepsilon} \subset \overline{\Omega_{3\varepsilon/4}} \subset \Omega_\varepsilon$. This function satisfies $\sigma_\lambda \phi_{\Omega;\varepsilon} = \phi_{\sigma_\lambda(\Omega);\varepsilon/\lambda}$ and $\tau_c \phi_{\Omega;\varepsilon} = \phi_{\tau_c(\Omega);\varepsilon}$. Now define

$$E_{\Omega;\varepsilon} u = \phi_{\Omega;\varepsilon} \mathcal{E}_\Omega u. \tag{2.70}$$

An application of part 2 then proves continuity. Since the extension operator commutes with translations and dilations we have

$$\tau_c E_{\Omega;\varepsilon} u = \tau_c (\phi_{\Omega;\varepsilon} \mathcal{E}_\Omega u) = (\tau_c \phi_{\Omega;\varepsilon}) \tau_c \mathcal{E}_\Omega u = \phi_{\tau_c(\Omega);\varepsilon} \mathcal{E}_{\tau_c(\Omega)} \tau_c u = E_{\tau_c(\Omega);\varepsilon} \tau_c u,$$

and

$$\sigma_\lambda E_{\Omega;\varepsilon} u = \sigma_\lambda (\phi_{\Omega;\varepsilon} \mathcal{E}_\Omega u) = \phi_{\sigma_\lambda(\Omega);\varepsilon/\lambda} \sigma_\lambda \mathcal{E}_\Omega u = \phi_{\sigma_\lambda(\Omega);\varepsilon/\lambda} \mathcal{E}_{\sigma_\lambda(\Omega)} \sigma_\lambda u = E_{\sigma_\lambda(\Omega);\varepsilon/\lambda} \sigma_\lambda u,$$

as required.

Part 4 Abbreviate the extension operator to $\mathcal{E} = \mathcal{E}_\Omega$. By part 5 Remark 82, if $u \in H_s^m(\Omega)$ then $\mathcal{E}u \in H_s^m$. But the norm $|||\cdot|||_{m,2}$ is equivalent to the Fourier transform norm

$$\int \prod_{k=1}^{d} \left(1 + \xi_k^2\right)^m |\widehat{u}(\xi)|^2 \, d\xi,$$

and so for $\alpha < m$

$$\begin{aligned} \int \left|\xi^\alpha \widehat{\mathcal{E}u}\right| &= \int |\xi^\alpha| \prod_{k=1}^{d} \left(1 + \xi_k^2\right)^{-\frac{m}{2}} \prod_{k=1}^{d} \left(1 + \xi_k^2\right)^{\frac{m}{2}} \left|\widehat{\mathcal{E}u}(\xi)\right| d\xi \\ &\leq \int \prod_{k=1}^{d} \left(1 + \xi_k^2\right)^{-\frac{1}{2}} \prod_{k=1}^{d} \left(1 + \xi_k^2\right)^{\frac{m}{2}} \left|\widehat{\mathcal{E}u}(\xi)\right| d\xi. \end{aligned}$$

An application of the Cauchy-Schwartz inequality now proves the inequalities $\int \left|\xi^\alpha \widehat{\mathcal{E}u}\right| < \infty$ for $\alpha < m$. Thus $\widehat{\mathcal{E}u} \in L^1$ and the identity A.1: $|\xi|^{2(m-1)} = \sum\limits_{|\alpha|=m-1} \frac{(m-1)!}{\alpha!}\xi^{2\alpha}$, implies $|\xi|^{m-1}\widehat{\mathcal{E}u} \in L^1$. Hence by part 2 Lemma 13 we have $\mathcal{E}u \in C_B^{(m-1)}$ and so $u \in C_B^{(m-1)}(\Omega)$. ■

We will need some more properties of the operator $\delta_2^{2l\mathbf{1}}$ which was introduced in Theorem 77.

Lemma 84 *Some properties of $\delta_2^{2l\mathbf{1}}$. Suppose w is the extended B-spline weight function with parameters n and l. Then:*

1. *$\delta_2^{2l\mathbf{1}} : H_s^n \to X_w^0$ is continuous.*
2. *If $f \in H_s^n$ and $\operatorname{supp} f \subset (-1,1)^d$ then $k_{l,d}\delta_2^{2l\mathbf{1}} f = f + \mathcal{A}_{2l} f$, where $\operatorname{supp} \mathcal{A}_{2l} f \subset \left([-1,1]^d\right)^c$ and $k_{l,d} = (-1)^{ld}\binom{2l}{l}^{-d}$.*
3. *$k_{l,d}\delta_2^{2l\mathbf{1}} : H_{s0}^n\left((-1,1)^d\right) \to X_w^0$ is a continuous extension mapping.*

Proof. Part 1 If $f \in H_s^n$ then $\widehat{\delta_2^{2l\mathbf{1}} f} = (2i\sin\xi_k)^{2l\mathbf{1}}\widehat{f}$ and

$$\left\|\delta_2^{2l\mathbf{1}} f\right\|_{w,0}^2 = \int \frac{\xi^{2n\mathbf{1}}}{(\sin\xi_k)^{2l\mathbf{1}}}\left|(2i\sin\xi_k)^{2l\mathbf{1}}\widehat{f}\right|^2 d\xi \le 2^{4ld}\int \xi^{2n\mathbf{1}}\left|\widehat{f}\right|^2 d\xi = 2^{4ld}\left\|D^{n\mathbf{1}} f\right\|_2^2$$
$$< 2^{4ld}\left|\left\|f\right\|\right|_{n,2}^2.$$

Part 2 From 2.58

$$\begin{aligned} k_{l,d}\delta_2^{2l\mathbf{1}} f &= k_{l,d}\sum_{\beta\le 2l}(-1)^{|\beta|}\binom{2l\mathbf{1}}{\beta}\tau_{2\beta-2l\mathbf{1}} f \\ &= k_{l,d}(-1)^{|l\mathbf{1}|}\binom{2l\mathbf{1}}{l\mathbf{1}} f + k_{l,d}\sum_{\beta\le 2l,\beta\ne l\mathbf{1}}(-1)^{|\beta|}\binom{2l\mathbf{1}}{\beta}\tau_{2\beta-2l\mathbf{1}} f \\ &= f + k_{l,d}\sum_{\beta\le 2l,\beta\ne l\mathbf{1}}(-1)^{|\beta|}\binom{2l\mathbf{1}}{\beta}\tau_{2\beta-2l\mathbf{1}} f. \end{aligned}$$

Now $\operatorname{supp} f \subset (-1,1)^d$ implies $\operatorname{supp}\tau_{2\beta-2l\mathbf{1}} f \subset \left([-1,1]^d\right)^c$ when $\beta \ne l\mathbf{1}$, and so $\operatorname{supp}\mathcal{A}_{2l} f \subset \left([-1,1]^d\right)^c$.

Part 3 Follows directly from part 2. ■

The next result allows us to use the local space $H_s^n(\Omega)$ as a source of data functions when the interpolation data is contained in a bounded open set Ω. This substantially simplifies the generation of data functions in comparison to the approach used in Subsection 2.6.1 for $n = 1$ but calculating an extension of the data function and its norm may need some work.

Theorem 85 *Let w be an extended B-spline weight function on $\mathbb{R}^d$ with parameters n and l. Suppose Ω is an open, bounded set which permits one of the extension operators $\mathcal{E}_\Omega$ described in parts 5 to 7 of Remark 82. Now if C is any open cube containing $\overline{\Omega}$ and $0 < \varepsilon < \operatorname{dist}(\Omega, C^c)$ then there exists a continuous linear mapping $E : H_s^n(\Omega) \to X_w^0$ and a linear mapping $\mathcal{B} : H_s^n(\Omega) \to X_w^0$ such that:*

1. *$E = E_{\Omega;\varepsilon} + \mathcal{B}$, where $E_{\Omega;\varepsilon} : H_s^n(\Omega) \to H_s^n$ is the extension mapping defined by 2.70,*
2. *$\operatorname{supp}\mathcal{B}f \subset \overline{C}^{\,c}$ and $\operatorname{supp} E_{\Omega;\varepsilon} f \subset \Omega_\varepsilon$ and,*
3. *E is an extension mapping.*
4. *$E = k_{l,d}\sum\limits_{\beta\le 2l}(-1)^\beta\binom{2l\mathbf{1}}{\beta}\tau_{\frac{2\beta-2l}{m}} E_{\Omega;\varepsilon}$, where $k_{l,d} = (-1)^{ld}\binom{2l}{l}^{-d}$ and,*

5. $\|Ef\|_{w,0} \le |k_{l,d}| (2m)^{2ld} \left\| D^{n\mathbf{1}} E_{\Omega;\varepsilon} f \right\|_2$.

Proof. Parts 1, 2, 3 The relevant commutative diagram is 2.73. Let c be the centre of the cube C and choose an integer $m > 0$ such that $(C - c)/m \subset [-1,1]^d$. We have $\sigma_m \tau_c(\Omega) = (\Omega - c)/m \subset [-1,1]^d$. Set $\Omega' = (\Omega - c)/m$. Choose $f \in H_s^n(\Omega)$. By part 3 Lemma 83, $\sigma_{1/m}\tau_{-c} f \in H_s^n(\Omega')$ can be extended by the operator $E_{\Omega',\varepsilon/m}$ to $\mathbb{R}^d$ as a function $g \in H_{s0}^n\left((-1,1)^d\right)$ which has support in $\Omega'_{\varepsilon/m}$. Lemma 84 therefore implies $k_{l,d}\delta_2^{2l\mathbf{1}} g$ is an extension of $\sigma_{1/m}\tau_{-c} f$ to X_w^0. Part 2 Theorem 77 now implies that $\tau_c \sigma_m k_{l,d} \delta_2^{2l\mathbf{1}} g \in X_w^0$. Set

$$Ef = \tau_c \sigma_m k_{l,d} \delta_2^{2l\mathbf{1}} E_{\Omega',\varepsilon/m} \sigma_{1/m} \tau_{-c} f. \tag{2.71}$$

Since $0 < \varepsilon/m \le \operatorname{dist}\left(\frac{\Omega - c}{m}, \left((-1,1)^d\right)^c\right)$, part 3 Lemma 83 and then part2 Lemma 84 allow us to write

$$\begin{aligned} Ef &= \tau_c \sigma_m k_{l,d} \delta_2^{2l\mathbf{1}} \sigma_{1/m} \tau_{-c} E_{\Omega;\varepsilon} f \\ &= \tau_c \sigma_m (I + \mathcal{A}_{2l}) \sigma_{1/m} \tau_{-c} E_{\Omega;\varepsilon} f \\ &= E_{\Omega;\varepsilon} f + \tau_c \sigma_m \mathcal{A}_{2l} \sigma_{1/m} \tau_{-c} E_{\Omega;\varepsilon} f. \end{aligned} \tag{2.72}$$

Now set

$$\mathcal{B} = \tau_c \sigma_m \mathcal{A}_{2l} \sigma_{1/m} \tau_{-c} E_{\Omega;\varepsilon},$$

so that $\operatorname{supp} \mathcal{B} f \subset \tau_{-c}\sigma_{1/m}\left(\operatorname{supp} \mathcal{A}_{2l} \sigma_{1/m} \tau_{-c} E_{\Omega;\varepsilon} f\right) \subset \tau_{-c}\sigma_{1/m}\left(\left([-1,1]^d\right)^c\right)$
$\subset \overline{C}^{\,c}$ and thus E is an extension.

Regarding continuity consider 2.71. By part 1 Lemma 83 the operator $\sigma_{1/m}\tau_{-c} : H_s^n(\Omega) \to H_s^n(\Omega')$ is continuous, the extension operator from $H_s^n(\Omega')$ to
$H_{s0}^n\left((-1,1)^d\right)$ is continuous by part 3 Lemma 83, $\delta_2^{2l\mathbf{1}} g : H_{s0}^n\left((-1,1)^d\right) \to X_w^0$ is continuous by Lemma 84, and $\tau_c \sigma_m : X_w^0 \to X_w^0$ is continuous by part 2 Theorem 77.

$$\begin{array}{ccc} H_s^n(\Omega) & \xrightarrow[E_{\Omega';\varepsilon/m}\sigma_{1/m}\tau_{-c}]{} & H_{s0}^n\left((-1,1)^d\right) \\ exten \downarrow & & exten \downarrow k_{l,d}\delta_2^{2l\mathbf{1}} \\ X_w^0 & \xleftarrow{\tau_c\sigma_m} & X_w^0 \end{array} \tag{2.73}$$

Part 4 Use equation 2.58 to calculate $\delta_2^{l\mathbf{1}}$ and then substitute into 2.72 and compose the translations and dilations.

Part 5 If l is even we start with 2.79. Using the estimates 2.61 and 2.59, for $f \in H_s^n(\Omega)$

$$\begin{aligned} \|Ef\|_{w,0} = \left\| \tau_c \sigma_m k_{l,d} \delta_2^{l\mathbf{1}} \sigma_{1/m} \tau_{-c} E_{\Omega;\varepsilon} f \right\|_{w,0} &= k_{l,d} \left\| \sigma_m \delta_2^{l\mathbf{1}} \sigma_{1/m} \tau_{-c} E_{\Omega;\varepsilon} f \right\|_{w,0} \\ &\le k_{l,d} m^{(l-n+1/2)d} \left\| \delta_2^{l\mathbf{1}} \sigma_{1/m} \tau_{-c} E_{\Omega;\varepsilon} f \right\|_{w,0} \\ &= k_{l,d} m^{(l-n+1/2)d} 2^{ld} \left\| D^{n\mathbf{1}} \sigma_{1/m} \tau_{-c} E_{\Omega;\varepsilon} f \right\|_2 \\ &= k_{l,d} (2m)^{ld} \left\| D^{n\mathbf{1}} \tau_{-c} E_{\Omega;\varepsilon} f \right\|_2 \\ &= k_{l,d} (2m)^{ld} \left\| D^{n\mathbf{1}} E_{\Omega;\varepsilon} f \right\|_2. \end{aligned}$$

∎

Definition 86 ***The restriction space*** $X_w^0(\Omega)$

Suppose that w is a weight function with property W2 and suppose $\Omega \subset \mathbb{R}^d$ is a open set.

Define the vector space $X_w^0(\Omega) = \left\{ r_\Omega u : u \in X_w^0 \right\}$, where $r_\Omega u$ denotes the action of the linear operator which restricts u to Ω. Endow $X_w^0(\Omega)$ with the norm

$$\|v\|_{w,0,\Omega} = \inf\left\{ \|u\|_{w,0} : u \in X_w^0,\ r_\Omega u = v \right\}. \tag{2.74}$$

Clearly $\|r_\Omega u\|_{w,0,\Omega} \le \|u\|_{w,0}$ so $r_\Omega : X_w^0 \to X_w^0(\Omega)$ is continuous. This implies $r_\Omega^{-1} v$ is closed and since $\|v\|_{w,0,\Omega} = \operatorname{dist}\left(0, r_\Omega^{-1} v\right)$ it follows that $\|v\|_{w,0,\Omega} = \|u\|_{w,0}$ for a unique $u \in X_w^0$ and $\|v\|_{w,0,\Omega}$ is actually well-defined and a norm.

Remark 87 *The dilation* $c^{-d/2}\sigma_c : X_w^0(\Omega) \to X_{\sigma_c w}^0(\sigma_c\Omega)$ *is an isometric homeomorphism.*

Corollary 88 *Let w be an extended B-spline weight function on $\mathbb{R}^d$ with parameters n and l. Suppose Ω is an open, bounded set which permits one of the extension operators $\mathcal{E}_\Omega$ described in parts 5 to 7 of Remark 82.*

Then $H_s^n(\Omega) \hookrightarrow X_w^0(\Omega)$ and this inclusion is valid for any scaled extended B-spline weight function with appropriately scaled Ω.

Proof. In Theorem 85 a continuous extension $E : H_s^n \hookrightarrow X_w^0$ was constructed for the extended B-spline weight functions w so that by Definition 86, $r_\Omega E : H_s^n(\Omega) \to X_w^0(\Omega)$ is a continuous inclusion. An application of Remark 87 now proves this corollary for scaled extended B-splines. ■

Observe that the local space $X_w^0(\Omega)$ can be regarded as the set of data functions for the minimal norm interpolation problem where the data is contained in Ω, and Corollary 88 proves that $H_s^n(\Omega) \subset X_w^0(\Omega)$ when w is a scaled extended B-spline weight function with parameters $n \le l$. The minimum norm interpolation problem can be defined in terms of extensions of these spaces to $\mathbb{R}^d$ and the matrix equation for the interpolant only uses values of the data function on Ω. **Thus $H_s^n(\Omega)$ can be used as a source of data functions for the interpolation problem when w is a scaled extended B-spline weight function**.

The last corollary will now be strengthened to characterize the data functions for the minimal norm interpolation problem for the scaled extended B-splines.

Lemma 89 *Suppose w is a scaled extended B-spline weight function with parameters n, l, and $\Omega \subset \mathbb{R}^d$ is a bounded open set. Then $X_w^0(\Omega) \hookrightarrow H_s^n(\Omega)$.*

Proof. Assume first that w is an unscaled extended B-spline weight function. Suppose $v \in X_w^0(\Omega)$ and let v^e be any extension of v to X_w^0. Noting that $n \le l$, if $\alpha \le n$ we have

$$\begin{aligned}
\left\|v^e\right\|_{w,0} = \int w\left|\widehat{v^e}\right|^2 = \int \frac{s^{2n\mathbf{1}}}{(\sin s_i)^{2l\mathbf{1}}}\left|\widehat{v^e}(s)\right|^2 ds &= \int s^{2\alpha}\frac{s^{2(n\mathbf{1}-\alpha)}}{(\sin s_i)^{2l\mathbf{1}}}\left|\widehat{v^e}(s)\right|^2 ds \\
&\ge \int s^{2\alpha}\frac{s^{2(n\mathbf{1}-\alpha)}}{(\sin s_i)^{2(n\mathbf{1}-\alpha)}}\left|\widehat{v^e}(s)\right|^2 ds \\
&\ge \int_\Omega s^{2\alpha}\left|\widehat{v^e}(s)\right|^2 ds \\
&= \int_\Omega \left|\widehat{D^\alpha v^e}(s)\right|^2 ds \\
&= \left\|D^\alpha v\right\|_{2,\Omega}^2,
\end{aligned}$$

and so $D^\alpha v \in L^2(\Omega)$ for $\alpha \le n$, which means $v \in H_s^n(\Omega)$. Further, the definition of the norm on $X_w^0(\Omega)$ implies $\left\|D^\alpha v\right\|_{2,\Omega} \le \left\|v^e\right\|_{w,0,\Omega}$ when $\alpha \le n$, verifying that the inclusion of $H_s^n(\Omega)$ in $X_w^0(\Omega)$ is continuous.

An application of Remark 87 now proves this lemma for scaled extended B-splines. ■

Combining Lemma 89 and Corollary 88 we obtain our local characterization of the interpolation data sets:

Theorem 90 *Suppose w is a scaled extended B-spline weight function with parameters n, l and $\Omega \subset \mathbb{R}^d$ is open, bounded set which permits an extension operator $\mathcal{E}_\Omega$ described in parts 5 to 7 of Remark 82.*

Then $H_s^n(\Omega) = X_w^0(\Omega)$ as sets and their norms are equivalent.

2.7.1 Extensions of Theorem90

In Theorem 90 it was shown that $H_s^n(\Omega) = X_w^0(\Omega)$ where w is a scaled extended B-spline weight function with parameters n, l and $\Omega \subset \mathbb{R}^d$ is open, bounded and satisfies the uniform cone condition e.g. §4.4 Adams [1]. Here we will extend the class of weight functions to which this result applies, although it still essentially only applies to tensor product weight functions.

The operator $\delta_2^{2l\mathbf{1}}$ has the property that $\widehat{\delta_2^{2l\mathbf{1}}u}(\xi) = (\sin\xi_k)^{2l\mathbf{1}}\,\widehat{u}(\xi)$ so that when multiplied by the weight function it's zeros can be used to cancel out the weight function poles, as in part 1 of Lemma 84,

leaving a function of polynomial increase of at most n. This ensures the continuity of $\delta_2^{2l\mathbf{1}} : H_s^n \to X_w^0$; so $(\sin \xi_k)^{l\mathbf{1}}$ corresponds to g below and motivates condition 2.75. It was also shown in Lemma 84 that $f \in H_{s0}^n\left((-1,1)^d\right)$ implies $k_{l,d}\delta_2^{2l\mathbf{1}} f = f + \mathcal{A}_{2l} f$, where $\operatorname{supp} \mathcal{A}_{2l} f \subset \left([-1,1]^d\right)^c$, and this prompted 2.76, where $\delta_2^{2l\mathbf{1}}$ corresponds to $g(D)$.

Suppose g is a **bounded function** for which there exist constants $C_\alpha \geq 0$ such that

$$|g(\xi)|^2 w(\xi) \leq \sum_{\alpha \leq n} C_\alpha \xi^{2\alpha}. \tag{2.75}$$

When $u \in H_s^n$ we have $u \in L^2 \subset S'$ and $\widehat{u} \in L^2 \subset L_{loc}^1$ and so boundedness of g implies the pseudodifferential operator $g(D) = (g\widehat{u})^\vee$ maps H_s^n to X_w^0 continuously with $\|g(D)u\|_{w,0} = \sqrt{\int w |g|^2 |\widehat{u}|^2} \leq \max_\alpha \sqrt{C_\alpha} \, \||u|\|_{n,2}$. Now assume that

$$\operatorname{supp}(g(D)f - f) \subset \left([-1,1]^d\right)^c, \quad f \in H_{s0}^n\left((-1,1)^d\right), \tag{2.76}$$

which means that $g(D) : H_{s0}^n\left((-1,1)^d\right) \to X_w^0$ is a continuous extension. The commutative diagram for the construction of the extension $H_s^n(\Omega) \to X_w^0$ is now 2.77 cf. diagram 2.73.

$$\begin{array}{ccc}
H_s^n(\Omega) & \underset{E_{\Omega';\varepsilon/m}\sigma_{1/m}\tau_{-c}}{- \longrightarrow} & H_{s0}^n\left((-1,1)^d\right) \\
exten \;\Big\downarrow & & exten \;\Big\downarrow\; g(D) \\
X_w^0 & \overset{\tau_c\sigma_m}{\longleftarrow -} & X_w^0
\end{array} \tag{2.77}$$

From the diagram we see that $\sigma_m : X_w^0 \to X_w^0$ must be continuous where $m > 0$ is an integer. This was proven for the spline weight functions in Theorem 77 and inspection of this proof shows that continuity still holds if

$$w(\xi/m) \leq c_m w(\xi), \qquad \xi \in \mathbb{R}^d, \tag{2.78}$$

for some constant $c_m > 0$. However, m must be chosen so that $(C-c)/m \subset [-1,1]^d$ where c is the centre of a cube containing $\overline{\Omega}$. Thus we have the following extension to Theorem 85:

Theorem 91 *Suppose:*

1. *w is a weight function on $\mathbb{R}^d$.*

2. *Ω is an open, bounded set which permits an extension operator $\mathcal{E}_\Omega$ described in parts 5 to 7 of Remark 82 and let C be any open cube with center c containing $\overline{\Omega}$. Set $0 < \varepsilon < \operatorname{dist}(\Omega, C^c)$.*

3. *there exists a bounded function g with properties 2.75 and 2.76.*

4. *there exists an integer $m > 0$ for which 2.78 holds and $(C-c)/m \subset [-1,1]^d$.*

Then there exists a continuous extension mapping $E : H_s^n(\Omega) \to X_w^0$ and a continuous mapping $\mathcal{B} : H_s^n(\Omega) \to X_w^0$ such that:

1. *$E = E_{\Omega;\varepsilon} + \mathcal{B}$, where $E_{\Omega;\varepsilon} : H_s^n(\Omega) \to H_s^n$ is the extension mapping defined by 2.70,*

2. *$\operatorname{supp} E_{\Omega;\varepsilon} f \subset \Omega_\varepsilon$ and $\operatorname{supp} \mathcal{B} f \subset \overline{C}^{\,c}$.*

3. *$E = k_{l,d} \sum_{\beta \leq 2l} (-1)^\beta \binom{2l\mathbf{1}}{\beta} \tau_{(2\beta - 2l)/m} E_{\Omega;\varepsilon}$, where $k_{l,d} = (-1)^{ld} \binom{2l}{l}^{-d}$ and,*

4. *$\|Ef\|_{w,0} \leq |k_{l,d}| (2m)^{2ld} \left\|D^{n\mathbf{1}} E_{\Omega;\varepsilon} f\right\|_2$.*

We now have an extended version of Corollary 88.

Corollary 92 *Let w be a weight function and Ω be an open set with the properties described in Theorem 91.*

Then $H_s^n(\Omega) \hookrightarrow X_w^0(\Omega)$ and this inclusion also valid for any scaled weight function with appropriately scaled Ω.

Proof. From Theorem 91 there is a continuous extension $E : H_s^n \hookrightarrow X_w^0$ so that by Definition 86, $r_\Omega E : H_s^n(\Omega) \to X_w^0(\Omega)$ is a continuous inclusion. An application of Remark 87 now proves this corollary for scaled weight functions. ■

The next result characterizes the data functions locally.

Corollary 93 *Suppose the assumptions of Theorem 91 hold and we also assume that there exist constant $c_\alpha > 0$ such that*

$$w(\xi) \geq c_\alpha \xi^{2\alpha}, \qquad \xi \in \Omega \text{ and } \alpha \leq n. \tag{2.79}$$

Then $X_w^0(\Omega) = H_s^n(\Omega)$ as sets and their norms are equivalent.

Proof. Inspection of Lemma 89 reveals that it still holds when the weight function satisfies property 2.79. Thus $X_w^0(\Omega) \hookrightarrow H_s^n(\Omega)$ and by Corollary 92, $X_w^0(\Omega) = H_s^n(\Omega)$ as sets with equivalent norms. ■

Remark 94 *If the inequalities 2.75 and 2.79 in Theorem 91 and Corollary 93 are replaced by ones in which the multi-index has range defined by $|\alpha| \leq n$ instead of $\alpha \leq n$ then both the theorem and the corollary are still valid but with $H_s^n(\Omega)$ replaced by the Sobolev space $H^n(\Omega)$ and $H_{s0}^n(\Omega)$ replaced by $H_0^n(\Omega)$.*

The Sobolev splines 1.15 are a case in point: here $w(\xi) = \left(1 + |\xi|^2\right)^n$. If we set $g = 1$ then by the identity of part 6 Definition 221, conditions 2.75 and 2.79 are satisfied for $c_\alpha = C_\alpha = \binom{k}{\alpha}$, clearly 2.76 is true and finally 2.78 is true for $c_w = 1$, $\Omega = \mathbb{R}^d$. Thus $X_w^0(\Omega) = H^n(\Omega)$ which confirms the result of Example 1.3.3.

3

Central difference weight functions

3.1 Introduction

To understand the theory of interpolants and smoothers presented in this document it is not necessary to read this chapter which introduces the *central difference weight functions*. However, we note that Section 3.5 contains local data space results specifically designed for tensor product weight functions and that central difference weight functions are used as examples in the chapters dealing with smoothers.

This chapter introduces a large new class of weight functions based on central difference operators and it is shown that they are weight functions. The basis functions are calculated and smoothness estimates and upper and lower bounds are derived for the weight functions and the basis functions.

Like the B-splines, their data functions are characterized locally as Sobolev-like spaces.

Pointwise convergence results are derived for the basis function interpolant using results from Chapter 2. These give orders of convergence identical to those obtained for the extended B-splines in. This chapter presents no numerical interpolation experiments.

This is a large class of weight functions and I have only calculated a couple of basis functions. The calculations can be tedious so one really needs to be able to have theoretical results which enable you to tell beforehand that the basis function has some especially good properties regarding interpolation - I have no results to offer in this regard except Remark 112 which might mean that it may be possible to usefully characterize the basis functions or characterize a large subset of the basis functions.

The central difference weight functions are defined as follows: Suppose $q \in L^1\left(\mathbb{R}^1\right)$, $q \neq 0$, $q(\xi) \geq 0$ and $l \geq n \geq 1$ are integers. The univariate *central difference weight function* with parameters n, l is defined by

$$w(\xi) = \frac{\xi^{2n}}{\Delta_{2l}\widehat{q}(\xi)}, \quad \xi \in \mathbb{R}^1,$$

where Δ_{2l} is central difference operator

$$\Delta_{2l} f(\xi) = \sum_{k=-l}^{l} (-1)^k \binom{2l}{k+l} f(-k\xi), \quad l = 1, 2, 3, \ldots, \ \xi \in \mathbb{R}^1.$$

For example, $\Delta_2 f(\xi) = -(f(\xi) - 2f(0) + f(-\xi))$. The multivariate central difference weight function is defined by tensor product.

It is shown in Theorem 98 that w belongs to the class of zero order weight functions introduced in Chapter 1 for some κ iff $\int\limits_{|\xi| \geq R} |\xi|^{2n-1} q(\xi)\, d\xi < \infty$ for some $R \geq 0$. Here κ satisfies $\kappa + 1/2 < n$. The central difference weight functions are closely related to the *extended B-splines* defined by 1.20 and a discussion of their genesis is given in Subsection 3.2.1.

Various bounds are derived for w. For example, in Corollary 109 it is shown that if $\int_{|t|\geq R} t^{2l} q(t)\,dt < \infty$ then for any $r > 0$ there exist constants $c_r, c_r', k_r, k_r' > 0$ such that

$$k_r \xi^{2n} \leq w(\xi) \leq k_r' \xi^{2n}, \quad |\xi| \geq r,$$
$$\frac{c_r}{\xi^{2(l-n)}} \leq w(\xi) \leq \frac{c_r'}{\xi^{2(l-n)}}, \quad |\xi| \leq r.$$

By Theorems 111 and 115 the univariate *central difference basis function* is

$$G(s) = (-1)^{(l-n)} \int\limits_{\mathbb{R}^1} \left(D^{2(l-n)} (*\Lambda)^l\right) \left(\frac{s}{t}\right) |t|^{2n-1} q(t)\,dt, \quad s \in \mathbb{R}^1,$$

with $G \in C_B^{(2n-2)}$ and $D^{2n-1}G$ bounded in $C^{(2n-1)}(\mathbb{R}\backslash 0)$. The multivariate basis function is defined as a tensor product. For $k \leq 2n-2$, $D^k G$ is uniformly Lipschitz continuous of order 1 (Theorem 118).

In Section 3.5 we will characterize the *data function space* locally as a Sobolev-like space 5. A localization result specifically for tensor product weight functions is derived in Theorem 123. This supplies important information about the data functions and makes it easy to choose data functions for numerical experiments.

Results for the pointwise convergence of the interpolant to its data function on a bounded data domain are derived using results from Chapter 1. These results are summarized in Table 3.1 and give *pointwise orders of convergence* identical to those obtained for the extended B-splines.

This chapter does not present the results of any numerical experiments.

3.2 Central difference weight functions

To introduce the central difference weight functions we will need some central difference operators:

Definition 95 ***The univariate central difference operators*** δ_ξ *and* Δ_{2l} *on* $\mathbb{R}^1$

For a continuous function f we define the central difference operators

$$\delta_\xi f(x) = f\left(x + \frac{\xi}{2}\right) - f\left(x - \frac{\xi}{2}\right), \quad x, \xi \in \mathbb{R}^1.$$

and

$$\Delta_{2l} f(\xi) = \sum_{k=-l}^{l} (-1)^k \binom{2l}{k+l} f(-k\xi), \quad l = 1,2,3,\ldots, \; \xi \in \mathbb{R}^1. \tag{3.1}$$

This definition of Δ_{2l} ensures that $\Delta_{2l,\xi}\left(e^{-i\xi t}\right) \geq 0$, which is part 3 of the next lemma. For example:

$$\Delta_2 f(\xi) = -\left(f(\xi) - 2f(0) + f(-\xi)\right),$$
$$\Delta_4 f(\xi) = f(2\xi) - 4f(\xi) + 6f(0) - 4f(-\xi) + f(-2\xi).$$

Lemma 96 *The central difference operators of Definition 95 have the following properties:*

1. $\Delta_{2l} f(\xi) = \left(\delta_\xi^{2l} f\right)(0)$.
2. *Regarding monomials and polynomials:*
   (a) $\Delta_{2l}\left(\xi^{2m}\right) = \left(\sum\limits_{k=-l}^{l} (-1)^k \binom{2l}{k+l} k^{2m}\right) \xi^{2m}$ *when* $m \geq 0$;
   (b) $\sum\limits_{k=-l}^{l} (-1)^k \binom{2l}{k+l} k^{2m} = \begin{cases} 0, & if\ m < l, \\ (-1)^l (2l)!, & if\ m = l; \end{cases}$
   (c) $\Delta_{2l}\left(\xi^{2m+1}\right) = 0$;
   (d) $\Delta_{2l}(p(\xi)) = 0$ *when p is a polynomial with degree less than* $2l$;
   (e) $\Delta_{2l}\left(\xi^{2l}\right) = (-1)^l (2l)! \xi^{2l}$;
3. $\Delta_{2l,\xi}\left(e^{-i\xi t}\right) = 2^{2l} \sin^{2l}(\xi t/2)$. *Here $\Delta_{2l,\xi}$ indicates Δ_{2l} is acting on the variable ξ in* $e^{-i\xi t}$.

3.2.1 Motivation and definition

The derivation of the central difference weight function is based on the approximation of the reciprocal of the B-spline weight function w_s using a mollifier. From 1.20 we have for parameters n and l

$$\frac{1}{w_s(\xi)} = \frac{\sin^{2l}\xi}{\xi^{2n}}, \quad 1 \le n \le l.$$

Now $\sin^{2l}\xi = \left(\frac{e^{i2\xi}-e^{-i2\xi}}{2i}\right)^{2l}$ so that by part 3 Lemma 96 with $t=2$, $\frac{1}{w_s(\xi)} = \frac{1}{2^{2l}}\frac{\Delta_{2l}\left(e^{-i2\xi}\right)}{\xi^{2n}}$ or using the Fourier transform $F[\cdot](\xi)$

$$\frac{1}{w_s(\xi)} = \frac{(2\pi)^{1/2}}{2^{2l}}\frac{\Delta_{2l}F[\delta(x-2)]}{\xi^{2n}}.$$

We now want to approximate the delta function δ in the distribution sense using the sequence of mollifier $k\psi(kx)$ where $k=1,2,3,\ldots$, $\psi \in L^1(\mathbb{R}^1)$, $\psi \ge 0$ and $\int \psi = 1$. Since $k\psi(kx) \to \delta$ as $k \to \infty$ in the sense of distributions, $k\psi(k(x-2)) \to \delta(x-2)$ and so as $k \to \infty$

$$\frac{\sqrt{2\pi}}{2^{2l}}\frac{\Delta_{2l}F[k\psi(k(x-2))]}{\xi^{2n}} \to \frac{1}{w_s(\xi)} \ \ on\ \mathcal{D}'\left(\mathbb{R}^1\backslash 0\right).$$

Here we will not try here to extend the convergence to S' but observe that we can write

$$\frac{\sqrt{2\pi}}{2^{2l}}\frac{\Delta_{2l}F[k\psi(k(x-2))]}{\xi^{2n}} = \frac{\Delta_{2l}\widehat{q_k}}{\xi^{2n}},$$

where

$$q_k(x) = \frac{\sqrt{2\pi}}{2^{2l}}k\psi(k(x-2)) \in L^1\left(\mathbb{R}^1\right) \ \ and\ q_k \ge 0.$$

Thus $\frac{\Delta_{2l}\widehat{q_k}}{\xi^{2n}} \to \frac{1}{w_s(\xi)}$ on $\mathcal{D}'\left(\mathbb{R}^1\backslash 0\right)$ and if we are lucky $\frac{\xi^{2n}}{\Delta_{2l}\widehat{q_k}(\xi)}$ may be a weight function. Indeed, we use this idea to define a central difference weight function and justify calling it a weight function using the next theorem.

Later in Section 3.4 we will show that if an extra condition is imposed on the function ψ then $\frac{\Delta_{2l}\widehat{q_k}}{\xi^{2n}} \to \frac{1}{w_s(\xi)}$ in L^1 and as tempered distributions, and that the corresponding sequence of central difference basis functions converges uniformly pointwise to the extended B-spline basis function.

Definition 97 ***Univariate central difference weight functions***
Suppose that $q \in L^1\left(\mathbb{R}^1\right)$, $q \ne 0$, $q(\xi) \ge 0$ and $l, n \ge 0$ are integers. The univariate central difference weight function is defined by

$$w(\xi) = \frac{\xi^{2n}}{\Delta_{2l}\widehat{q}(\xi)}, \quad \xi \in \mathbb{R}^1. \tag{3.2}$$

In d dimensions the central difference weight function with parameters n and l is the tensor product of the univariate weight function.

The next result justifies these definitions.

Theorem 98 *Suppose w is the function on $\mathbb{R}^1$ introduced in Definition 97. Then w is an even function satisfying weight function property W1 for $\mathcal{A} = \{0\}$. Further, w satisfies property W2 for parameter κ iff l and n satisfy $\kappa + 1/2 < n \le l$ and q satisfies*

$$\int\limits_{|\xi|\ge R} |\xi|^{2n-1} q(\xi)\, d\xi < \infty, \quad for\ some\ R \ge 0. \tag{3.3}$$

Proof. The weight function w is an even function since it is clear from equation 3.1 that $\Delta_{2l}\widehat{q}$ is an even function. From part 5 of Lemma 96 we have $\Delta_{2l,\xi}e^{-i\xi t} = 2^{2l}\sin^{2l}(\xi t/2)$ so that

$$\begin{aligned}\frac{1}{w(\xi)} = \frac{\Delta_{2l}\widehat{q}(\xi)}{\xi^{2n}} &= \frac{1}{\sqrt{2\pi}}\frac{1}{\xi^{2n}}\int \left(\Delta_{2l,\xi}e^{-i\xi t}\right) q(t)\, dt \\ &= \frac{2^{2l}}{\sqrt{2\pi}}\frac{1}{\xi^{2n}}\int \sin^{2l}(\xi t/2)\, q(t)\, dt. \end{aligned}\tag{3.4}$$

From Theorem 98 we have $w(\xi) > 0$ when $\xi \neq 0$. Since $q \in L^1$ implies $\widehat{q} \in C_B^{(0)}$, from the definition of w we have $w \in C^{(0)}\left(\mathbb{R}^1 \backslash 0\right)$ and $w > 0$ on $\mathbb{R}^1 \backslash 0$.

Regarding property W2, choose $\lambda \geq 0$. Again using Theorem 98

$$\int \frac{\xi^{2\lambda} d\xi}{w(\xi)} = \frac{2^{2l}}{\sqrt{2\pi}} \int \frac{1}{\xi^{2(n-m)}} \int \sin^{2l}(\xi t/2)\, q(t)\, dt d\xi = \frac{2^{2l}}{\sqrt{2\pi}} \int \int \frac{\sin^{2l}(\xi t/2)}{\xi^{2(n-\lambda)}} d\xi\, q(t)\, dt.$$

where, because the integrand is non-negative, the integrals all exist iff the last integral exists. Now make the change of variables $\eta = \xi t/2$ so that

$$\begin{aligned} \int \frac{\xi^{2\lambda} d\xi}{w(\xi)} &= \frac{2^{2(l-n+\lambda)-1}}{\sqrt{2\pi}} \int \int \frac{\sin^{2l}(\eta)}{\eta^{2(n-\lambda)}} d\eta\, |t|^{2(n-\lambda)-1} q(t)\, dt \\ &= \frac{2^{2(l-n+\lambda)-1}}{\sqrt{2\pi}} \int \frac{\eta^{2\lambda}}{w_s(\eta)} d\eta \int |t|^{2(n-\lambda)-1} q(t)\, dt. \end{aligned} \tag{3.5}$$

Since w_s is an extended B-spline weight function the first integral of 3.5 exists iff $0 \leq \lambda \leq \kappa$ and $\kappa + 1/2 < n \leq l$. Further, since $q \in L^1$, $\int |t|^{2(n-\lambda)-1} q(t)\, dt$ exists iff $\int\limits_{|t| \geq R} |t|^{2(n-\lambda)-1} q(t)\, dt < \infty$ for some $R \geq 0$ iff $\int\limits_{|t| \geq R} |t|^{2(n-\lambda)-1} q(t)\, dt < \infty$ for $\lambda = 0$ and some $R \geq 0$. Thus it can be seen that w has property W2 for κ iff $\kappa + 1/2 < n \leq l$ and $\int\limits_{|t| \geq R} |t|^{2n-1} q(t)\, dt < \infty$. ■

The central difference weight function is related to the extended B-spline weight function as follows:

Theorem 99 *Suppose w_s is the extended B-spline weight function with parameters n, l. Suppose w_c is a central difference weight function with parameters $n, l, q(\cdot)$ and property 3.3. Then these weight functions are related by the equation*

$$\frac{1}{w_c(\xi)} = \frac{2^{2(l-n)}}{\sqrt{2\pi}} \int_{\mathbb{R}^1} \frac{1}{w_s(\xi t/2)} t^{2n} q(t)\, dt. \tag{3.6}$$

Proof. From equation 3.4 of the proof of Theorem 98

$$\begin{aligned} \frac{1}{w_c(\xi)} = \frac{1}{\sqrt{2\pi}} \frac{2^{2l}}{\xi^{2n}} \int \sin^{2l}(\xi t/2)\, q(t)\, dt &= \frac{2^{2(l-n)}}{\sqrt{2\pi}} \int \frac{\sin^{2l}(\xi t/2)}{(\xi t/2)^{2n}} t^{2n} q(t)\, dt \\ &= \frac{2^{2(l-n)}}{\sqrt{2\pi}} \int \frac{t^{2n} q(t)}{w_s(\xi t/2)} dt. \end{aligned}$$

■

3.2.2 Weight function bounds and smoothness

We will be mostly concerned with smoothness properties and deriving lower bounds for the weight function. It turns out that the weight function is continuous except at the origin where it may have a singularity.

The next theorem will give some properties of the central difference weight function if condition 3.3 is strengthened to 3.7.

Theorem 100 *Suppose w is a central difference weight function with the properties assumed in Theorem 98. Suppose in addition*

$$\int\limits_{|t| \geq R} t^{2n} q(t)\, dt < \infty. \tag{3.7}$$

Then $1/w \in C_B^{(0)}\left(\mathbb{R}^1\right)$ and

$$\lim_{\xi \to 0} \frac{1}{w(\xi)} = \begin{cases} 0, & if \quad n < l, \\ \frac{1}{\sqrt{2\pi}} \int t^{2n} q(t)\, dt, & if \quad n = l. \end{cases} \tag{3.8}$$

Also, there exists a positive constant C_1 such that

$$\frac{1}{w(\xi)} \leq \frac{C_1}{\xi^{2n}}, \quad \xi \neq 0. \tag{3.9}$$

Proof. From Theorem 98 we know that w is continuous and positive outside the origin. Thus $1/w$ is also continuous and positive outside the origin. We now show that $1/w$ is also continuous at the origin by proving the limits 3.8. The existence of the integral $\int\limits_{|t|\geq R} t^{2n} q(t)\,dt$ implies $t^{2n} q(t) \in L^1$ and so $\widehat{q} \in C_B^{(2n)}$ by part 2 of Lemma 13. Hence the Taylor series with integral remainder of Appendix A.7 can be used to expand each term of 3.1 about the origin up to a polynomial of degree $2n-1$. By part 3 of Lemma 96 Δ^{2l} annihilates polynomials of degree less than $2l$ and so:

$$\begin{aligned}
\Delta^{2l}\widehat{q}(\xi) &= \sum_{k=-l}^{l} (-1)^{|k|} \tbinom{2l}{k+l} \widehat{q}(-k\xi) \\
&= \sum_{k=-l}^{l} (-1)^{|k|} \tbinom{2l}{k+l} \sum_{j<2n} \frac{\left(D^j \widehat{q}\right)(0)}{j!} (-k\xi)^j + \\
&+ \sum_{k=-l}^{l} (-1)^{|k|} \tbinom{2l}{k+l} \frac{(k\xi)^{2n}}{(2n-1)!} \int_0^1 s^{2n-1} \left(D^{2n}\widehat{q}\right)((s-1)k\xi)\,ds \\
&= \sum_{j<2n} \frac{\left(D^j \widehat{q}\right)(0)}{j!} - \xi^j \sum_{k=-l}^{l} (-1)^{|k|} \tbinom{2l}{k+l} k^j + \\
&+ \sum_{k=-l}^{l} (-1)^{|k|} \tbinom{2l}{k+l} \frac{(-k\xi)^{2n}}{(2n-1)!} \int_0^1 s^{2n-1} \left(D^{2n}\widehat{q}\right)((s-1)k\xi)\,ds \\
&= \sum_{k=-l}^{l} (-1)^{|k|} \tbinom{2l}{k+l} \frac{(k\xi)^{2n}}{(2n-1)!} \int_0^1 s^{2n-1} \left(D^{2n}\widehat{q}\right)((s-1)k\xi)\,ds \\
&= \frac{\xi^{2n}}{(2n-1)!} \sum_{k=-l}^{l} (-1)^{|k|} \tbinom{2l}{k+l} k^{2n} \int_0^1 s^{2n-1} \left(D^{2n}\widehat{q}\right)((s-1)k\xi)\,ds. \qquad (3.10)
\end{aligned}$$

Thus

$$\frac{1}{w(\xi)} = \frac{\Delta^{2l}\widehat{q}(\xi)}{\xi^{2n}} = \frac{1}{(2n-1)!} \sum_{k=-l}^{l} (-1)^{|k|} \tbinom{2l}{k+l} k^{2n} \int_0^1 s^{2n-1} \left(D^{2n}\widehat{q}\right)((s-1)k\xi)\,ds.$$

Because $D^{2n}\widehat{q} \in C_B^{(0)}$ we can apply the Lebesgue-dominated convergence theorem to the sequence $\left\{ s^{2n-1} \left(D^{2n}\widehat{q}\right) \left((1-s)\left(-\frac{k}{j}\right)\right) \right\}_{j=1}^{\infty}$ to obtain

$$\begin{aligned}
\lim_{\xi\to 0} \frac{1}{w(\xi)} &= \frac{1}{(2n-1)!} \sum_{k=-l}^{l} (-1)^{|k|} \tbinom{2l}{k+l} k^{2n} \int_0^1 s^{2n-1} \lim_{\xi\to 0} \left(D^{2n}\widehat{q}\right)((s-1)k\xi)\,ds \\
&= \frac{1}{2n!} \left(\sum_{k=-l}^{l} (-1)^{|k|} \tbinom{2l}{k+l} k^{2n} \right) \left(D^{2n}\widehat{q}\right)(0) \\
&= \begin{cases} 0, & if \quad n<l, \\ \frac{1}{\sqrt{2\pi}} \int t^{2l} q(t)\,dt, & if \quad n=l. \end{cases}
\end{aligned}$$

where the last step used property 4 of Lemma 96.

The next thing to prove is inequality 3.9. But from equation 3.4,

$$\frac{1}{w(\xi)} = \frac{1}{\sqrt{2\pi}} \frac{1}{\xi^{2n}} \int \sin^{2l}(\xi t/2)\, q(t)\,dt \leq \frac{1}{\sqrt{2\pi}} \left(\int q\right) \frac{1}{\xi^{2n}}.$$

Thus $1/w$ is bounded outside the unit sphere. Since $1/w$ is continuous everywhere it now follows that $1/w$ is bounded everywhere. ■

The next result places the generally stronger condition 3.11 on the function q in exchange for estimates of the weight function behavior near the origin.

Theorem 101 *Suppose w is a central difference weight function with parameters n, l, q which satisfy the conditions of Theorem 98, and also suppose that*

$$\int_{|t|\geq R} t^{2l} q(t)\, dt < \infty. \tag{3.11}$$

Then

$$\lim_{\xi\to 0} \frac{1}{w(\xi)\, \xi^{2(l-n)}} = \frac{1}{2\sqrt{2\pi}} \int_{\mathbb{R}^1} t^{2l} q(t)\, dt, \quad n \leq l, \tag{3.12}$$

and given $a > 0$ there exist constants $c_a, c_a' > 0$ such that

$$c_a \leq w(\xi)\, \xi^{2(l-n)} \leq c_a', \quad |\xi| \leq a. \tag{3.13}$$

Regarding smoothness: $w \in C^{(2l)}\left(\mathbb{R}^1 \backslash 0\right)$ and $w(\xi)\, \xi^{2(l-n)} \in C^{(0)}$.

Proof. Assumption 3.11 implies that $\widehat{q} \in C_B^{(2l)}$. Hence we can set $n = l$ in the proof of equation 3.10 of Theorem and obtain

$$\Delta^{2l}\widehat{q}(\xi) = \frac{\xi^{2l}}{(2l-1)!} \sum_{k=-l}^{l} (-1)^{|k|} \tbinom{2l}{k+l} k^{2l} \int_0^1 s^{2l-1} \left(D^{2l}\widehat{q}\right)((s-1)k\xi)\, ds,$$

or, since $\frac{1}{w(\xi)} = \xi^{-2n}\Delta^{2l}\widehat{q}(\xi)$ for $n \leq l$,

$$\frac{1}{w(\xi)\, \xi^{2(l-n)}} = \xi^{-2l}\Delta^{2l}\widehat{q}(\xi) = \frac{1}{(2l-1)!} \sum_{k=-l}^{l} (-1)^{|k|} \tbinom{2l}{k+l} k^{2l} \int_0^1 s^{2l-1} \left(D^{2l}\widehat{q}\right)((s-1)k\xi)\, ds.$$

Following the rest of Theorem 98 we obtain the limit 3.12.

To prove 3.13 note that since w is continuous, 3.12 implies that there exists $a_1 > 0$ such that

$$\left|\frac{1}{w(\xi)\, \xi^{2(l-n)}} - \frac{1}{\sqrt{2\pi}} \int t^{2l} q(t)\, dt\right| \leq \frac{1}{2\sqrt{2\pi}} \int t^{2l} q(t)\, dt, \quad |\xi| \leq a_1,$$

and so

$$\frac{1}{2\sqrt{2\pi}} \int t^{2l} q(t)\, dt \leq \frac{1}{w(\xi)\, \xi^{2(l-n)}} \leq \frac{3}{2\sqrt{2\pi}} \int t^{2l} q(t)\, dt, \quad |\xi| \leq a_1,$$

or

$$\frac{1}{\frac{3}{2\sqrt{2\pi}} \int t^{2l} q(t)\, dt} \leq w(\xi)\, \xi^{2(l-n)} \leq \frac{1}{\frac{1}{2\sqrt{2\pi}} \int t^{2l} q(t)\, dt}.$$

Now choose

$$c_a = \frac{1}{\frac{3}{2\sqrt{2\pi}} \int t^{2l} q(t)\, dt}, \quad c_a' = 3c_a.$$

If $a \leq a_1$ we are done. If $a > a_1$ we use the continuity of w and 3.9 to prove that there exist $C_1, C_2 > 0$ such that $0 < C_1 \leq w(\xi) \leq C_2$ for $a_1 \leq |\xi| \leq a$. The inequalities 3.13 then follow easily.

Regarding smoothness, we noted at the start of this proof that $\widehat{q} \in C_B^{(2l)}$ and consequently $\Delta^{2l}\widehat{q_k} \in C_B^{(2l)}$. Further, equation 3.4 implies that $\Delta_{2l}\widehat{q}(\xi) > 0$ when $\xi \neq 0$. But $w(\xi) = \frac{\xi^{2n}}{\Delta^{2l}\widehat{q}(\xi)}$ and so $w \in C^{(2l)}(\mathbb{R}\backslash 0)$. Finally, 3.12 of this theorem implies that $\lim\limits_{\xi\to 0} \frac{1}{w(\xi)\xi^{2(l-n)}}$ exists and is positive. Thus $\lim\limits_{\xi\to 0} w(\xi)\, \xi^{2(l-n)}$ exists and so the function $w(\xi)\, \xi^{2(l-n)}$ is continuous. ■

A multi-dimensional analogue of Theorem 98 is:

Corollary 102 *Suppose the functions $\{q_j\}_{j=1}^{d}$ satisfy $q_j \in L^1\left(\mathbb{R}^1\right)$, $q_j \neq 0$, $q_j(x) \geq 0$. Define the tensor product function $w(\xi) = \prod\limits_{j=1}^{d} \frac{\xi_j^{2n}}{\Delta^{2l}\widehat{q_j}(\xi_j)}$ for integers $l, n \geq 0$.*

*Then w is a weight function on $\mathbb{R}^d$. We call it a tensor product **central difference** weight function. Further, w satisfies property W2 for parameter κ iff l and n also satisfy the conditions of Theorem 98 i.e. $\kappa + 1/2 < n \leq l$ and $\int_{|t|\geq R} |t|^{2n-1} q_j(t)\, dt < \infty$ for some R_j and $1 \leq j \leq d$.*

Proof. That w satisfies property W2 for κ since the tensor product of univariate weight functions having property W2 for κ also satisfies property W2 for κ. This is part 2 of Theorem 4. ■

Remark 103 *The weight functions of Corollary 102 can be further generalized by choosing different values of n and l for each dimension.*

The next result gives some weight function properties which follow directly from Theorem 101.

Corollary 104 *Suppose w is a central difference weight function with parameters n, l, q which satisfy the conditions of Theorem 101. Suppose in addition that $l = n$. Then w satisfies:*

1. $w \in C^{(0)}\left(\mathbb{R}^1\right) \cap C^{(2l)}\left(\mathbb{R}^1 \backslash 0\right)$.

   There exist constants $c_1, c_2 > 0$ such that:

2. $w(\xi) \geq c_1$ *for all ξ and,*
3. $w(\xi) \geq c_2 \xi^{2l}$ *for all ξ.*
4. $\lim\limits_{\xi \to 0} w(\xi) = \left(\frac{1}{2^{2n}\sqrt{2\pi}} \int t^{2n} q(t)\, dt\right)^{-1}$.

3.2.3 Upper bounds for the weight function

To derive upper bounds on central difference weight functions we first derive the weight function for the case when q is a rectangular function.

Theorem 105 *Suppose $b > a$ and q_R is the rectangular function*

$$q_R(\xi) = \begin{cases} 1, & \xi \in [a,b], \\ 0, & \xi \notin [a,b]. \end{cases} \tag{3.14}$$

Then $q = q_R$ satisfies the conditions of Theorem 98 for $n = l = 1$ and the corresponding weight function is

$$w_R(\xi) = \frac{\sqrt{2\pi}}{8} \frac{\xi^2}{\frac{b-a}{2} - \frac{\cos\left(\frac{b+a}{2}\xi\right)\sin\left(\frac{b-a}{2}\xi\right)}{\xi}}, \tag{3.15}$$

with bounds

$$\frac{\sqrt{2\pi}}{6(b-a)}\xi^2 \leq w_R(\xi) \leq \frac{\sqrt{2\pi}}{2(b-a)}\xi^2, \quad |\xi| \geq \frac{4}{b-a}. \tag{3.16}$$

Proof. From 3.4

$$\frac{1}{w_R(\xi)} = \frac{4}{\sqrt{2\pi}} \frac{1}{\xi^2} \int_{\mathbb{R}^1} \left(\sin^2 \frac{\xi t}{2}\right) q_R(t)\, dt = \frac{4}{\sqrt{2\pi}} \frac{1}{\xi^2} \int_a^b \sin^2 \frac{\xi t}{2} dt = \frac{8}{\sqrt{2\pi}} \frac{1}{\xi^2} \int_a^b \sin^2 \frac{\xi t}{2} dt.$$

The integral is an even function of ξ so now assume $\xi > 0$. We now use the change of variables $s = \xi t/2$, $ds = (\xi/2)\, dt$ with new range $[a\xi/2, b\xi/2]$. Hence

$$\begin{aligned}
\int_a^b \sin^2 \frac{\xi t}{2} dt = \frac{2}{\xi} \int_{a\xi/2}^{b\xi/2} \sin^2 s ds &= \frac{2}{\xi} \left[\frac{s}{2} - \frac{\sin 2s}{4}\right]_{a\xi/2}^{b\xi/2} \\
&= \frac{2}{\xi} \left(\frac{b-a}{4}\xi - \frac{\sin b\xi - \sin a\xi}{4}\right) \\
&= \frac{b-a}{2} - \frac{\sin b\xi - \sin a\xi}{2\xi} \\
&= \frac{b-a}{2} - \frac{1}{\xi} \cos\left(\frac{b+a}{2}\xi\right) \sin\left(\frac{b-a}{2}\xi\right),
\end{aligned}$$

so that

$$\frac{1}{w_R(\xi)} = \frac{8}{\sqrt{2\pi}} \frac{1}{\xi^2} \left(\frac{b-a}{2} - \frac{1}{\xi} \cos\left(\frac{b+a}{2}\xi\right) \sin\left(\frac{b-a}{2}\xi\right)\right),$$

as required.

Regarding the inequality 3.16, if $|\xi| \geq \frac{4}{b-a}$ then $\frac{1}{|\xi|} \leq \frac{b-a}{4}$ and

$$w_R(\xi) = \frac{\sqrt{2\pi}}{8} \frac{\xi^2}{\frac{b-a}{2} - \frac{\cos\left(\frac{b+a}{2}\xi\right)\sin\left(\frac{b-a}{2}\xi\right)}{\xi}} \leq \frac{\sqrt{2\pi}}{8} \frac{\xi^2}{\frac{b-a}{2} - \frac{1}{|\xi|}} \leq \frac{\sqrt{2\pi}}{8} \frac{\xi^2}{(b-a)/4}$$
$$= \frac{\sqrt{2\pi}}{2(b-a)}\xi^2.$$

Also, since $\frac{1}{|\xi|} \leq \frac{b-a}{4}$

$$w_R(\xi) = \frac{\sqrt{2\pi}}{8} \frac{\xi^2}{\frac{b-a}{2} - \frac{\cos\left(\frac{b+a}{2}\xi\right)\sin\left(\frac{b-a}{2}\xi\right)}{\xi}} \geq \frac{\sqrt{2\pi}}{8} \frac{\xi^2}{\frac{b-a}{2} + \frac{1}{|\xi|}} \geq \frac{\sqrt{2\pi}}{6(b-a)}\xi^2.$$

■

Corollary 106 *The weight function w_R of Theorem 105 is a $C_{BP}^{\infty}\left(\mathbb{R}^1\right)$ function.*

Proof. We continue w_R to $\mathbb{C}^1$ as the function of a complex variable:

$$w_R(z) = \frac{\sqrt{2\pi}}{4} \frac{1}{\frac{(b-a)z - (\sin bz - \sin az)}{z^3}}, \quad z \in \mathbb{C}^1,$$

and note that

$$\frac{(b-a)z - (\sin bz - \sin az)}{z^3} = \frac{\left(b^3 - a^3\right)}{3!} - \frac{\left(b^5 - a^5\right)z^2}{5!} + \frac{\left(b^7 - a^7\right)z^4}{7!} - \ldots,$$

is analytic and non-zero in a neighborhood of the origin. The theory of *division of a power series* implies that $w_R(z)$ is also analytic in a neighborhood of the origin. Thus w_R is C^∞ in a neighborhood of the origin.

To prove that w_R is C^∞ away from the origin we write

$$w_R(\xi) = \frac{\sqrt{2\pi}}{8} \frac{1}{\left(\frac{b-a}{2}\right)^3} f_R\left(\frac{b-a}{2}\xi\right),$$

where

$$f_R(\xi) = \frac{\xi^2}{1 - \cos\left(\frac{b+a}{b-a}\xi\right)\frac{\sin\xi}{\xi}}.$$

It is clear that $\cos\left(\frac{b+a}{b-a}\xi\right)\frac{\sin\xi}{\xi} = 1$ iff $\xi = 0$, so that $f_R \in C^\infty$ away from the origin and hence that $w \in C^\infty$ away from the origin. Finally, since we can write $|f_R(\xi)| \leq \xi^2$ we have $w \in C_{BP}^\infty$. ■

The previous theorem will now be applied to derive an upper bound for the central difference weight functions with parameters $n = l = 1$.

Corollary 107 *For the case $n = l = 1$ the central difference weight function $w(\xi) = \frac{\xi^2}{\Delta_2 \widehat{q}(\xi)}$ of Theorem 98 satisfies an inequality of the form*

$$w(\xi) \leq c w_R(\xi), \quad \xi \in \mathbb{R}^1.$$

where w_R was defined in Theorem 105.

Proof. By assumption $q \in L^1\left(\mathbb{R}^1\right)$, $q(\xi) \geq 0$ for all $\xi \in \mathbb{R}^1$. Therefore, since an integrable function can be defined using simple functions i.e. the characteristic functions of intervals, there exist constants $b > a$ and $c > 0$ such that $q(\xi) \geq c q_R(\xi)$ for $\xi \in [a, b]$ where q_R be the rectangular function defined by 3.14. Then from 3.4 and Theorem 105

$$\frac{1}{w(\xi)} = \frac{4}{\sqrt{2\pi}} \frac{1}{\xi^2} \int\limits_{\mathbb{R}^1} \sin^2\frac{\xi t}{2} q(t)\,dt \geq \frac{4}{\sqrt{2\pi}} \frac{c}{\xi^2} \int_a^b \sin^2\frac{\xi t}{2} q_R(t)\,dt = \frac{c}{w_R(\xi)}.$$

■

For an arbitrary central difference weight function inequality 3.9 gave an order $2n$ lower bound on the growth of a weight function at infinity. We now derive an upper bound at infinity which is also of order $2n$.

Theorem 108 *There exist constants $0 \le a < b$ such that the central difference weight function $w(\xi) = \frac{\xi^{2n}}{\Delta_{2l}\widehat{q}(\xi)}$ of Theorem 98 satisfies inequalities of the form*

$$w(\xi) \le \begin{cases} \frac{c_2}{\xi^{2(l-n)}}, & |\xi| \le \frac{\pi}{b-a}, \\ 2c_1\xi^{2n}, & |\xi| > \frac{\pi}{b-a}. \end{cases} \tag{3.17}$$

where c_1 and c_2 are given by equations 3.21 and 3.23 respectively.

Proof. The proof will involve calculating a lower bound for $1/w$. Choose $0 \le a < b$ and c_0 such that $w(\xi) \ge c_0$ when $\xi \in [a,b]$. Set $\delta = b-a$. We start with equation 3.4 so that

$$\frac{1}{\xi^{2(l-n)}w(\xi)} = \frac{2^{2l}}{\sqrt{2\pi}}\frac{1}{\xi^{2l}}\int_{\mathbb{R}^1}\sin^{2l}\frac{\xi t}{2}q(t)\,dt \ge \frac{2^{2l}c_0}{\sqrt{2\pi}}\frac{1}{\xi^{2l}}\int_a^b \sin^{2l}\frac{\xi t}{2}dt.$$

Since w is an even function we can assume $\xi > 0$. Applying the change of variables $s = \xi t/2$ gives

$$\frac{1}{\xi^{2(l-n)}w(\xi)} \ge \frac{2^{2l+1}c_0}{\sqrt{2\pi}}\frac{1}{\xi^{2l+1}}\int_{a\xi/2}^{b\xi/2}\sin^{2l} s ds. \tag{3.18}$$

Now $b\xi/2 - a\xi/2 = \delta\xi/2 = k\pi + \sigma\pi$ for some $0 \le \sigma < 1$, and

$$\begin{aligned}\int_{a\xi/2}^{b\xi/2}\sin^{2l} sd &= \int_{a\xi/2}^{a\xi/2+k\pi}\sin^{2l} sds + \int_{a\xi/2+k\pi}^{a\xi/2+k\pi+\sigma\pi}\sin^{2l} sds \\ &= \int_0^{k\pi}\sin^{2l} sds + \int_{a\xi/2}^{a\xi/2+\sigma\pi}\sin^{2l} sds \\ &= 2k\int_0^{\pi/2}\sin^{2l} sds + \int_{a\xi/2}^{a\xi/2+\sigma\pi}\sin^{2l} sds.\end{aligned}$$

Since $\delta\xi/2 = k\pi + \sigma\pi$ implies $\delta\xi/2 < (k+1)\pi$ and $\frac{\delta\xi}{2\pi} - 1 < k$,

$$\int_{a\xi/2}^{b\xi/2}\sin^{2l} sd \ge 2k\int_0^{\pi/2}\sin^{2l} sds \ge \left(\frac{\delta\xi}{\pi} - 2\right)\int_0^{\pi/2}\sin^{2l} sds. \tag{3.19}$$

We now split the range of ξ into the intervals: $0 < \xi \le \frac{\pi}{\delta}$, $\frac{\pi}{\delta} \le \xi < \frac{4\pi}{\delta}$ and $\xi \ge \frac{4\pi}{\delta}$ and consider each as a separate case.

$\boxed{\text{Case 1: } \xi \ge \frac{4\pi}{\delta}}$ Here $k \ge 2$ and $\frac{2}{\xi} \le \frac{\delta}{2\pi}$ so that from 3.18 and 3.19

$$\begin{aligned}\frac{1}{\xi^{2(l-n)}w(\xi)} \ge \frac{2^{2l+1}c_0}{\sqrt{2\pi}}\frac{1}{\xi^{2l+1}}\int_{a\xi/2}^{b\xi/2}\sin^{2l} sds &\ge \frac{2^{2l+1}c_0}{\sqrt{2\pi}}\frac{1}{\xi^{2l+1}}\left(\frac{\delta\xi}{\pi} - 2\right)\int_0^{\pi/2}\sin^{2l} sds \\ &= \frac{2^{2l+1}c_0}{\sqrt{2\pi}}\frac{1}{\xi^{2l}}\left(\frac{\delta}{\pi} - \frac{2}{\xi}\right)\int_0^{\pi/2}\sin^{2l} sds \\ &\ge \frac{2^{2l+1}c_0}{\sqrt{2\pi}}\frac{1}{\xi^{2l}}\left(\frac{\delta}{\pi} - \frac{\delta}{2\pi}\right)\int_0^{\pi/2}\sin^{2l} sds \\ &= \frac{2^{2l+1}c_0}{\sqrt{2\pi}}\frac{1}{\xi^{2l}}\frac{\delta}{2\pi}\int_0^{\pi/2}\sin^{2l} sds \\ &= \frac{2^{2l}}{\sqrt{2\pi}}\frac{1}{\pi}\left(\int_0^{\pi/2}\sin^{2l} sds\right)\frac{\delta c_0}{\xi^{2l}} \\ &= \frac{2^{2l}}{\sqrt{2\pi}}\frac{1}{2}\frac{(2l)!}{2^{2l}(l!)^2}\frac{\delta c_0}{\xi^{2l}} \\ &= \frac{1}{2\sqrt{2\pi}}\frac{(2l)!}{(l!)^2}\frac{\delta c_0}{\xi^{2l}},\end{aligned}$$

and thus

$$w(\xi) \leq c_1 \xi^{2n}, \quad \xi \geq \frac{4\pi}{\delta}, \tag{3.20}$$

where

$$c_1 = 2\sqrt{2\pi}\frac{(l!)^2}{(2l)!}\frac{1}{(b-a)\,c_0}. \tag{3.21}$$

$\boxed{\text{Case 2: } 0 < \xi < \frac{\pi}{\delta}}$ Here $k = 0$ and $b\xi/2 - a\xi/2 = \delta\xi/2 < \pi/2$ so that $\sin s \geq \frac{2}{\pi}s$ and

$$\int_{a\xi/2}^{b\xi/2} \sin^{2l} s ds \geq \int_{a\xi/2}^{b\xi/2} \left(\frac{2}{\pi}s\right)^{2l} ds \geq \left(\frac{2}{\pi}\right)^{2l} \left[\frac{s^{2l+1}}{2l+1}\right]_{a\xi/2}^{b\xi/2}$$
$$= \frac{1}{2}\left(\frac{1}{\pi}\right)^{2l} \frac{1}{2l+1}\left(b^{2l+1} - a^{2l+1}\right)\xi^{2l+1}.$$

Thus by 3.18

$$\frac{1}{\xi^{2(l-n)} w(\xi)} \geq \frac{2^{2l+1}c_0}{\sqrt{2\pi}} \frac{1}{\xi^{2l+1}} \int_{a\xi/2}^{b\xi/2} \sin^{2l} s ds \geq \frac{2^{2l+1}c_0}{\sqrt{2\pi}} \frac{1}{2}\left(\frac{1}{\pi}\right)^{2l} \frac{1}{2l+1}\left(b^{2l+1} - a^{2l+1}\right)$$
$$\geq \frac{1}{\sqrt{2\pi}}\left(\frac{2}{\pi}\right)^{2l} \frac{b^{2l+1} - a^{2l+1}}{2l+1} c_0,$$

and

$$w(\xi) \leq \frac{c_2}{\xi^{2(l-n)}}, \quad 0 < \xi < \frac{\pi}{\delta}, \tag{3.22}$$

where

$$c_2 = \frac{1}{\sqrt{2\pi}}\left(\frac{2}{\pi}\right)^{2l} \frac{\left(b^{2l+1} - a^{2l+1}\right) c_0}{2l+1}. \tag{3.23}$$

$\boxed{\text{Case 3: } \frac{\pi}{\delta} \leq \xi < \frac{4\pi}{\delta}}$ Here

$$\int_{a\xi/2}^{b\xi/2} \sin^{2l} s ds \geq \int_0^{\pi/2} \sin^{2l} s ds,$$

so that by 3.18

$$\frac{1}{\xi^{2(l-n)} w(\xi)} \geq \frac{2^{2l+1}c_0}{\sqrt{2\pi}} \frac{1}{\xi^{2l+1}} \int_{a\xi/2}^{b\xi/2} \sin^{2l} s ds \geq \frac{2^{2l+1}c_0}{\sqrt{2\pi}} \frac{1}{\xi^{2l+1}} \int_0^{\pi/2} \sin^{2l} s ds$$
$$= \frac{2^{2l+1}c_0}{\sqrt{2\pi}} \frac{1}{\xi^{2l+1}} \frac{\pi}{2} \frac{(2l)!}{2^{2l}(l!)^2}$$
$$= \frac{\sqrt{2\pi}}{2} \frac{(2l)!}{(l!)^2} \frac{c_0}{\xi^{2l+1}}.$$

Thus

$$w(\xi) \leq \frac{2}{\sqrt{2\pi}} \frac{(l!)^2}{(2l)!} \frac{\xi^{2n+1}}{c_0} \leq \frac{2}{\sqrt{2\pi}} \frac{(l!)^2}{(2l)!} \frac{\xi^{2n}}{c_0} \frac{4\pi}{\delta} = 4\sqrt{2\pi}\frac{(l!)^2}{(2l)!} \frac{1}{c_0\delta}\xi^{2n} = 2c_1\xi^{2n},$$

and

$$w(\xi) \leq 2c_1\xi^{2n}, \quad \frac{\pi}{\delta} \leq \xi < \frac{4\pi}{\delta}. \tag{3.24}$$

Noting that w is an even function the inequalities 3.20, 3.22 and 3.24 imply 3.17 as desired. ■

We have the following results for central difference weight function near zero and infinity:

Corollary 109 *Suppose w is a central difference weight function on $\mathbb{R}^1$ with parameters $n, l, q(\cdot)$ which satisfies the condition 3.11. Then for any $r > 0$ there exist constants $c_r, c'_r, k_r, k'_r > 0$ such that*

$$k_r\xi^{2n} \leq w(\xi) \leq k'_r\xi^{2n}, \quad |\xi| \geq r, \tag{3.25}$$

and

$$\frac{c_r}{\xi^{2(l-n)}} \leq w(\xi) \leq \frac{c'_r}{\xi^{2(l-n)}}, \quad |\xi| \leq r. \tag{3.26}$$

Proof. Inequalities 3.25 follow from inequality 3.17 of Theorem 108 and inequality 3.9 of Theorem 100. Inequalities 3.26 are the inequalities 3.13 of Theorem 101. ■

Later we will be interested in the sequence of central difference weight functions $\{w_k\}$ which corresponds to $b = k^{-1} > 0$ and $a = -k^{-1}$ where k is a positive integer. These weight functions have the following properties:

Corollary 110 *Suppose $k \geq 1$ is an integer and R is the unit rectangular function with support on $[-1,1] \subset \mathbb{R}^1$.*

Then $q_k(\xi) = \frac{k}{2} R(k\xi)$ satisfies the conditions of Theorem 98 for $n = l = 1$ and the corresponding central difference weight function is

$$w_k(\xi) = \frac{\sqrt{2\pi}}{16} k^3 f_R\left(\frac{\xi}{k}\right),$$

where

$$f_R(\xi) = \frac{\xi^2}{1 - \frac{\sin\xi}{\xi}}.$$

Proof. The function $\frac{k}{2} R(k\xi)$ is a rectangular function equal to $\frac{k}{2}$ on $\left[-\frac{1}{k}, \frac{1}{k}\right]$. Thus $a = -\frac{1}{k}$ and $b = \frac{1}{k}$ so that by Theorem 105

$$w_k(\xi) = \frac{\sqrt{2\pi}}{16} \frac{\xi^2}{\frac{1}{k} - \frac{\sin(\xi/k)}{\xi}} = \frac{\sqrt{2\pi}}{16} \frac{\xi^3}{\frac{\xi}{k} - \sin\frac{\xi}{k}} = \frac{\sqrt{2\pi}}{16} k^3 f_R(\xi/k).$$

■

3.3 The central difference basis functions

We will now determine the basis functions of zero order generated by the central difference weight functions of Definition 97. In Theorem 29 it was shown that a univariate extended B-spline basis function G_s is such that $G_s \in C_0^{(2n-2)}\left(\mathbb{R}^1\right)$ and $D^{2n-1} G_s$ is a piecewise constant function. In the next two theorems we will prove analogues of these results for the central difference basis functions. In fact we show that a central difference basis function G_c is such that $G_c \in C_B^{(2n-2)}$, $D^{2n-1} G_c \in C^{(0)}\left(\mathbb{R}^1 \backslash 0\right)$ and $D^{2n-1} G_c$ is bounded a.e.

Theorem 111 *Suppose w is the tensor product central difference weight function defined in Corollary 102 and set*

$$G_\Lambda = (-1)^{(l-n)} D^{2(l-n)} (*\Lambda)^l. \tag{3.27}$$

Then this weight function generates the tensor product basis function $G(x) = \prod_{i=1}^{d} G_1(x_i)$ where

$$G_1(s) = \int_{\mathbb{R}^1} G_\Lambda\left(\frac{s}{t}\right) |t|^{2n-1} q(t)\, dt, \quad s \in \mathbb{R}^1, \tag{3.28}$$

$G_1 \in C_B^{(2n-2)}$ and G_1 is an even function. Further, for $1 \leq k \leq 2n-2$

$$D^k G_1(s) = \int_{\mathbb{R}^1} (\operatorname{sgn} t)^k \left(D^k G_\Lambda\right)\left(\frac{s}{t}\right) |t|^{2n-k-1} q(t)\, dt, \quad s \in \mathbb{R}^1, \tag{3.29}$$

Finally, for $0 \leq \tau \leq k$

$$\left|D^k G_1(s)\right| \leq \frac{\left\||x|^\tau D^k G_\Lambda\right\|_\infty \left\||t|^\tau q\right\|_1}{|s|^\tau}, \quad s \in \mathbb{R}^1 \backslash 0.$$

Here G_Λ is given by 3.27.

Proof. We first note that from Subsection 1.4.4 the multivariate basis function will just be the product of the univariate basis functions.

Because of the multiplicity of variables involved we will use the notation $F[\,]$ for the Fourier transform. We want to express $1/w$ as a Fourier transform. Continuing on from equation 3.4

$$\begin{aligned}\frac{1}{w(\xi)} = \tfrac{1}{\sqrt{2\pi}}\frac{1}{\xi^{2n}}\int_{\mathbb{R}^1} 2^{2l}\sin^{2l}\frac{\xi t}{2}q(t)\,dt &= \tfrac{1}{\sqrt{2\pi}}\int (\xi t)^{2(l-n)}\frac{\sin^{2l}\frac{\xi t}{2}}{(\xi t/2)^{2l}}t^{2n}q(t)\,dt \\ &= \int F\left[(-1)^{(l-n)}D^{2(l-n)}(*\Lambda)^l\right](\xi t)\,t^{2n}q(t)\,dt \\ &= \int F_x\left[G_\Lambda\left(\frac{x}{t}\right)\right](\xi)\,|t|^{2n-1}q(t)\,dt.\end{aligned}$$

Now

$$\begin{aligned}G_1(s) &= \tfrac{1}{\sqrt{2\pi}}\int \frac{e^{-i\xi s}}{w(\xi)}d\xi \\ &= \tfrac{1}{\sqrt{2\pi}}\int\int e^{-i\xi s}F_s\left[G_\Lambda\left(\tfrac{s}{t}\right)\right](\xi)\,|t|^{2n-1}q(t)\,dtd\xi,\end{aligned} \tag{3.30}$$

and we want to use Fubini's theorem to change the order of integration. Indeed

$$\begin{aligned}\int\int\left|e^{-i\xi s}F_s\left[G_\Lambda\left(\tfrac{s}{t}\right)\right](\xi)\,|t|^{2n-1}q(t)\right|d\xi dt &\le \int\int\left|F_s\left[G_\Lambda\left(\tfrac{s}{t}\right)\right](\xi)\right|d\xi\ \ |t|^{2n-1}q(t)\,dt \\ &= \int\int|F[G_\Lambda](\xi t)|\ \ |t|\,d\xi\ \ |t|^{2n-1}q(t)\,dt \\ &= \tfrac{1}{\sqrt{2\pi}}\int\int\frac{\sin^{2l}(\xi t/2)}{(\xi t/2)^{2n}}|t|\,d\xi\ \ |t|^{2n-1}q(t)\,dt \\ &= \tfrac{1}{\sqrt{2\pi}}\int\int\frac{\sin^{2l}(\eta/2)}{(\eta/2)^{2n}}d\eta\ \ |t|^{2n-1}q(t)\,dt \\ &= \tfrac{1}{\sqrt{2\pi}}\int\int F[G_\Lambda](\eta)\,d\eta\ \ |t|^{2n-1}q(t)\,dt \\ &= G_\Lambda(0)\int|t|^{2n-1}q(t)\,dt,\end{aligned}$$

so the order of integration can be reversed to give

$$G_1(s) = \tfrac{1}{\sqrt{2\pi}}\int\int e^{-i\xi s}F_s\left[G_\Lambda\left(\tfrac{s}{t}\right)\right](\xi)\,d\xi\ \ |t|^{2n-1}q(t)\,dt = \int G_\Lambda\left(\tfrac{s}{t}\right)|t|^{2n-1}q(t)\,dt,$$

as required.

By Theorem 98 w is even. Hence G_1 is even since w is real-valued. Also by Theorem 98, $\kappa+1/2<n$ so that $\lfloor 2\kappa\rfloor \le 2n-2$. But by Theorem 24 $G_1\in C_B^{(\lfloor 2\kappa\rfloor)}$ and so $G_1\in C_B^{(2n-2)}$.

To prove 3.29 we start with the equations 3.30 for G_1. If $k\le 2n-2$

$$\begin{aligned}D^kG_1(s) = \tfrac{1}{\sqrt{2\pi}}\int\frac{(-i\xi)^k e^{-i\xi s}}{w(\xi)}d\xi &= \tfrac{1}{\sqrt{2\pi}}\int\int e^{-i\xi s}(-i\xi)^k F_s\left[G_\Lambda\left(\tfrac{s}{t}\right)\right](\xi)\,|t|^{2n-1}q(t)\,dtd\xi \\ &= \tfrac{1}{\sqrt{2\pi}}\int\int e^{-i\xi s}F_s\left[\left(D^kG_\Lambda\right)\left(\tfrac{s}{t}\right)\right](\xi)\,t^{-k}|t|^{2n-1}q(t)\,dtd\xi.\end{aligned}$$

An argument similar to that used to change the order of integration in the case $k=0$ allows us to write

$$\begin{aligned}D^kG_1(s) &= \tfrac{1}{\sqrt{2\pi}}\int\int e^{-i\xi s}F_s\left[\left(D^kG_\Lambda\right)\left(\tfrac{s}{t}\right)\right](\xi)\,d\xi\ t^{-k}|t|^{2n-1}q(t)\,dt \\ &= \int\left(D^kG_\Lambda\right)\left(\tfrac{s}{t}\right)t^{-k}|t|^{2n-1}q(t)\,dt \\ &= \int(\operatorname{sgn}t)^k\left(D^kG_\Lambda\right)\left(\tfrac{s}{t}\right)|t|^{2n-k-1}q(t)\,dt,\end{aligned}$$

as required.
Again from 3.29, since $D^k G_\Lambda$ is bounded with bounded support, for any $0 \leq \tau \leq k$

$$\begin{aligned} |s|^\tau \left|D^k G_1(s)\right| &\leq \int_{\mathbb{R}^1} \left||s|^\tau \left(D^k G_\Lambda\right)\left(\frac{s}{t}\right)\right| |t|^{2n-k-1} q(t)\,dt \\ &= \int \left|\left|\frac{s}{t}\right|^\tau \left(D^k G_\Lambda\right)\left(\frac{s}{t}\right)\right| |t|^{2n-k-1+\tau} q(t)\,dt \\ &\leq \left\||x|^\tau D^k G_\Lambda\right\|_\infty \int |t|^{2n-k-1+\tau} q(t)\,dt \\ &= \left\||x|^\tau D^k G_\Lambda\right\|_\infty \left\||t|^{2n-k-1+\tau} q\right\|_1, \end{aligned}$$

so that

$$\left|D^k G_1(s)\right| \leq \frac{\left\||x|^\tau D^k G_\Lambda\right\|_\infty \left\||t|^{2n-k-1+\tau} q\right\|_1}{|s|^\tau}$$

■

Remark 112 *The expressions 3.29 and 3.28 are **multiplicative convolutions** and exponential substitutions would convert them into **additive convolutions**. Then, with appropriate modifications, Fourier transforms could be used to express q in terms of the basis function and perhaps even enable us to characterize the basis functions.*

However, we could instead use the theory of Mellin transforms and the fact that the Mellin transform of a multiplicative convolution is the product of the Mellin transforms.

Corollary 113 *Suppose G_1 is a one-dimensional central difference basis function generated by a weight function $w(\xi) = \frac{\xi^{2n}}{\Delta^{2l}\widehat{q}(\xi)}$ with property W2. Suppose also that q decreases faster than any power of $|\cdot|$ at infinity. Then for $0 \leq \tau$ and $0 \leq k \leq 2n-2$*

$$\left|D^k G_1(s)\right| \leq \frac{\left\||x|^\tau D^k G_\Lambda\right\|_\infty \left\||t|^\tau q\right\|_1}{|s|^\tau}, \quad s \in \mathbb{R}^1 \backslash 0.$$

For the next theorem we will prove the following version of Leibniz's theorem for differentiating under the integral sign.

Lemma 114 *(Leibniz) Suppose we have a function $f : \mathbb{R}^2 \to \mathbb{C}$ and a point $s_0 \in \mathbb{R}$. Suppose f has the following properties :*

1. *In some neighborhood of $\mathcal{N}(s_0)$ of s_0 and for almost all t, $f(s,t) \in C^{(0)}$ and $D_1 f(s,t)$ is piecewise continuous.*

2. *For each $s \in \mathcal{N}(s_0)$, $f(s,\cdot) \in L^1$ and $D_1 f(s,\cdot) \in L^1$.*

3. *There exist constants $c, \varepsilon > 0$ such that*

$$\int |D_1 f(s,t) - D_1 f(s_0,t)|\,dt \leq c\,|s - s_0|, \quad |s - s_0| \leq \varepsilon. \tag{3.31}$$

Then we can conclude that in some neighborhood of s_0, $D_1 \int f(s,t)\,dt = \int (D_1 f)(s,t)\,dt$ and this is a continuous function.

Proof. The conditions of parts 1 and 2 imply that for each fixed t, $f(s,t)$ is absolutely continuous as a function of s in a neighborhood of s_0. This absolute continuity enables us to write

$$\int \left(\frac{f(s,t) - f(s_0,t)}{h} - (D_1 f)(s_0,t)\right) dt = \frac{1}{h} \int \int_{s_0}^{s} \left((D_1 f)(u,t) - (D_1 f)(s_0,t)\right) du\,dt,$$

where $h = s - s_0$. Condition 3 of this lemma allows us to use Fubini's theorem to change the order of integration in the following calculation. Assuming that $s > s_0$

$$\begin{aligned}\left|\int \left(\frac{f(s,t) - f(s_0,t)}{h} - (D_1 f)(s_0,t)\right) dt\right| &\leq \frac{1}{h}\int\int_{s_0}^{s} \left|(D_1 f)(u,t) - (D_1 f)(s_0,t)\right| du\, dt \\ &= \frac{1}{h}\int_{s_0}^{s}\int \left|(D_1 f)(u,t) - (D_1 f)(s_0,t)\right| dt\, du \\ &\leq \frac{c}{h}\int_{s_0}^{s} (u - s_0)\, du \\ &= \frac{c}{2}(s - s_0).\end{aligned}$$

The same estimate can be obtained when $s < s_0$. ■

We have shown that the basis function lies in $C_B^{(2n-2)}$. Now we will use the lemma to consider the derivatives of order $2n - 1$.

Theorem 115 *Suppose G_1 is a one-dimensional central difference basis function generated by a weight function $w(\xi) = \frac{\xi^{2n}}{\Delta^{2l}\widehat{q}(\xi)}$ satisfying property W2. Suppose also that q is bounded a.e. i.e. for some constant $c_q > 0$*

$$q(t) \leq c_q \quad a.e. \tag{3.32}$$

Then, if G_Λ is given by 3.27,

$$D^{2n-1}G_1(s) = \int\limits_{\mathbb{R}^1} \operatorname{sgn}(t)\left(D^{2n-1}G_\Lambda\right)\left(\frac{s}{t}\right) q(t)\, dt, \quad s \in \mathbb{R}^1\backslash 0, \tag{3.33}$$

$D^{2n-1}G_1 \in C^{(0)}\left(\mathbb{R}^1\backslash 0\right)$ and $D^{2n-1}G_1$ is bounded a.e.

Further, for $0 \leq \tau \leq 2n - 1$

$$\left|D^{2n-1}G_1(s)\right| \leq \frac{\left\|x^\tau D^{2n-1}G_\Lambda\right\|_\infty \left\|t^\tau q\right\|_1}{|s|^\tau}, \quad s \in \mathbb{R}^1\backslash 0.$$

Proof. We will define

$$f(s,t) = \left(D^{2n-2}G_\Lambda\right)\left(\frac{s}{t}\right)|t|\, q(t),$$

and verify that f satisfies the three properties of Lemma 114. First note that from Theorem 29, $G_\Lambda \in C_0^{(2n-2)}$ and $D^{2n-1}G_\Lambda$ is a piecewise constant function with bounded support. Also, since G_1 has weight function property W2, $q \in L^1$ and $\int_{|\cdot|\geq R} |t|^{2n-1} q(t)\, dt < \infty$ so that $\int |t|\, q(t)\, dt < \infty$.

Property 1 Thus for each $t \neq 0$, $f(s,t) \in C^{(0)}$ and since

$$D_1 f(s,t) = \operatorname{sgn}(t)\left(D^{2n-1}G_\Lambda\right)\left(\frac{s}{t}\right) q(t), \tag{3.34}$$

it follows that $D_1 f(\cdot,t)$ is piecewise continuous.

Property 2 Since $G_\Lambda \in C_0^{(2n-2)}$, $D^{2n-2}G_\Lambda$ is bounded and hence

$$\int |f(s,t)|\, dt \leq \left\|D^{2n-2}G_\Lambda\right\|_\infty \int |t|\, q(t)\, dt < \infty,$$

and

$$\int |D_1 f(s,t)|\, dt \leq \int \left|\left(D^{2n-1}G_\Lambda\right)\left(\frac{s}{t}\right)\right| q(t)\, dt \leq \left\|D^{2n-1}G_\Lambda\right\|_\infty \int q(t)\, dt < \infty.$$

Property 3 Assume $s_0 \neq 0$. Suppose $\operatorname{supp} D^{2n-1}G_\Lambda \subset B(0;r)$ and set $t_{\min} = \max\{|s_0|, |s|\}/r$. Then

$$\begin{aligned}\int |D_1 f(s,t) - D_1 f(s_0,t)|\, dt &= \int \left|\left(D^{2n-1}G_\Lambda\right)\left(\frac{s}{t}\right) - \left(D^{2n-1}G_\Lambda\right)\left(\frac{s_0}{t}\right)\right| q(t)\, dt \\ &= \int\limits_{|t|\geq t_{\min}} \left|\left(D^{2n-1}G_\Lambda\right)\left(\frac{s}{t}\right) - \left(D^{2n-1}G_\Lambda\right)\left(\frac{s_0}{t}\right)\right| q(t)\, dt.\end{aligned}$$

Since $D^{2n-1}G_\Lambda$ is a step function with bounded support it can be written as the sum of a finite number of rectangular functions, say $\{a_k\Pi_k\}_{k=1}^m$, so that

$$\begin{aligned}\int |D_1 f(s,t) - D_1 f(s_0,t)|\, dt &\leq \sum_{k=1}^m a_k \int\limits_{|t|\geq t_{\min}} \left|\Pi_k\left(\frac{s}{t}\right) - \Pi_k\left(\frac{s_0}{t}\right)\right| q(t)\, dt \\ &\leq \left\|D^{2n-1}G_\Lambda\right\|_\infty \sum_{k=1}^m \int\limits_{|t|\geq t_{\min}} \left|\Pi_k\left(\frac{s}{t}\right) - \Pi_k\left(\frac{s_0}{t}\right)\right| q(t)\, dt. \end{aligned} \tag{3.35}$$

Suppose Π_k has support $\left[x_1^{(k)}, x_2^{(k)}\right]$. Observing that

$$\{x : \Pi_k(x) - \Pi_k(x+h) \neq 0\} \subseteq \bigcup_{j=1}^2 \left\{x : \left|x - x_j^{(k)}\right| \leq |h|\right\},$$

we have

$$\begin{aligned}\left\{t : \Pi_k\left(\frac{s_0}{t}\right) \neq \Pi_k\left(\frac{s}{t}\right), \quad |t| \geq t_{\min}\right\} &\subseteq \bigcup_{j=1}^2 \left\{t : \left|\frac{s_0}{t} - x_j^{(k)}\right| \leq \left|\frac{s}{t} - \frac{s_0}{t}\right|, \quad |t| \geq t_{\min}\right\} \\ &= \bigcup_{j=1}^2 \left\{t : \left|\frac{s_0}{t} - x_j^{(k)}\right| \leq \frac{|s-s_0|}{|t|}, \quad |t| \geq t_{\min}\right\}.\end{aligned}$$

Next apply the further constraint $|s-s_0| < |s_0|/2$ so that $x_j^{(k)} = 0$ implies the corresponding set is empty. If $x_1, x_2 \neq 0$

$$\begin{aligned}\left\{t : \Pi_k\left(\frac{s_0}{t}\right) \neq \Pi_k\left(\frac{s}{t}\right), \quad |t| \geq t_{\min}\right\} &\subseteq \bigcup_{j=1}^2 \left\{t : \left|t - \frac{s_0}{x_j^{(k)}}\right| \leq \frac{|s-s_0|}{\left|x_j^{(k)}\right|}, \quad |t| \geq t_{\min}\right\} \\ &= I_1^{(k)} + I_2^{(k)},\end{aligned}$$

with $0 \notin I_j^{(k)}$. We can now conclude from equation 3.35 that when $|s-s_0| < |s_0|/2$

$$\int |D_1 f(s,t) - D_1 f(s_0,t)|\, dt \leq \left\|D^{2n-1}G_\Lambda\right\|_\infty \sum_{k=1}^m \int\limits_{I_1^{(k)} \cup I_2^{(k)}} q = \left\|D^{2n-1}G_\Lambda\right\|_\infty \int\limits_{J(s_0,s)} q,$$

where, $J(s_0,s) = \bigcup\limits_{k=1}^m \left(I_1^{(k)} \cup I_2^{(k)}\right) = \bigcup\limits_{k=1}^m [a_k - b_k|s-s_0|, a_k + b_k|s-s_0|]$ with $b_k > 0$, $a_k \neq 0$, and $0 \notin J(s_0,s)$.

Now the additional assumption 3.32 we made about q comes into play. In fact

$$\begin{aligned}\int\limits_{J(s_0,s)} q(t)\, dt = \sum_{k=1}^m \int\limits_{a_k - b_k|s-s_0|}^{a_k + b_k|s-s_0|} q \leq c_q \sum_{k=1}^m \int\limits_{a_n - b_n|s-s_0|}^{a_n + b_n|s-s_0|} dt &\leq c_q \sum_{k=1}^m 2b_k|s-s_0| \\ &= 2c_q\left(\sum_{k=1}^m b_k\right)|s-s_0|,\end{aligned}$$

so that

$$\int |D_1 f(s,t) - D_1 f(s_0,t)|\, dt \leq 2c_q\left(\sum_{k=1}^m b_k\right)|s-s_0|,$$

when $|s-s_0| < |s_0|/2$ and $s_0 \neq 0$. This proves property 3 and hence 3.33 and $D^{2n-1}G_1 \in C^{(0)}\left(\mathbb{R}^1 \backslash 0\right)$.

From 3.33

$$\left|D^{2n-1}G_1(s)\right| \le \int_{\mathbb{R}^1} \left|\left(D^{2n-1}G_\Lambda\right)\left(\frac{s}{t}\right)\right| q(t)\,dt$$
$$\le \left\|D^{2n-1}G_\Lambda\right\|_\infty \int q(t)\,dt < \infty \ a.e.$$

Again from 3.33, since $D^{2n-1}G_\Lambda$ is bounded with bounded support, for any $0 \le \tau \le 2n-1$

$$|s|^\tau \left|D^{2n-1}G_1(s)\right| \le \int \left||s|^\tau \left(D^{2n-1}G_\Lambda\right)\left(\frac{s}{t}\right)\right| q(t)\,dt = \int \left|\left|\frac{s}{t}\right|^\tau \left(D^{2n-1}G_\Lambda\right)\left(\frac{s}{t}\right)\right| |t|^\tau q(t)\,dt$$
$$\le \left\||x|^\tau \left(D^{2n-1}G_\Lambda\right)\right\|_\infty \int |t|^\tau q(t)\,dt$$
$$= \left\||x|^\tau D^{2n-1}G_\Lambda\right\|_\infty \left\||t|^\tau q\right\|_1 .$$

■

3.3.1 Examples

Example 116 *In this example we show that the univariate function $\Lambda(x)^2$ is a central difference basis function with weight function $w(\xi) = \frac{\sqrt{2\pi}}{4}\frac{\xi^2}{1-\frac{\sin\xi}{\xi}}$.*

Suppose $l = n = 1$ and $q(x) = R(x)$ where R is the rectangular function with support $[-1,1]$ and height 1. In this case $G_\Lambda = \Lambda$ i.e. the hat function. Then the central difference basis function is given by

$$G(x) = \int_{-1}^{1} \Lambda\left(\frac{x}{t}\right)|t|\,R(t)\,dt = 2\int_0^1 \Lambda\left(\frac{x}{t}\right)tdt,$$

so that $G(x) = 0$ when $|x| \ge 1$, since $\Lambda(x) = 0$ when $|x| \ge 1$.
When $|x| < 1$

$$G(x) = 2\int_{|x|}^{1}\left(1-\frac{|x|}{t}\right)tdt = 2\int_{|x|}^{1}(t-|x|)\,dt = \left[t^2 - 2|x|\,t\right]_{|x|}^{1} = (1-2|x|) - \left(|x|^2 - 2|x|^2\right)$$
$$= 1 - 2|x| + |x|^2$$
$$= (1-|x|)^2 .$$

Thus $G(x) = \Lambda(x)^2$.
Since $\widehat{q}(\xi) = \frac{2}{\sqrt{2\pi}}\frac{\sin\xi}{\xi}$, by the definition of w

$$\frac{1}{w(\xi)} = \frac{\Delta_{2l}\widehat{q}(\xi)}{\xi^{2n}} = \frac{\Delta_2\widehat{q}(\xi)}{\xi^2} = \frac{-\left(\widehat{q}(\xi) - 2\widehat{q}(0) + \widehat{q}(-\xi)\right)}{\xi^2} = \frac{2}{\sqrt{2\pi}}\frac{2-2\frac{\sin\xi}{\xi}}{\xi^2}$$
$$= \frac{4}{\sqrt{2\pi}}\frac{\xi - \sin\xi}{\xi^3},$$

and thus $w(\xi) = \frac{\sqrt{2\pi}}{4}\frac{\xi^2}{1-\frac{\sin\xi}{\xi}}$.

Example 117 *Noting the convergence claims in the sentence before Definition 97 we will now construct a sequence of central difference basis functions which converges uniformly pointwise to the hat basis function. See Theorem 122 for a more general result. Here we choose the sequence*

$$q_k(t) = \frac{k}{2}R(kt), \quad k = 1,2,3,\ldots,\ t \in \mathbb{R}^1.$$

where R is the unit rectangular function with support $[-1,1]$. Suppose $l = n = 1$ so that the central difference basis function is:

$$G_k(x) = \int \Lambda\left(\frac{x}{t}\right)|t|\,q_k(t-1)\,dt = \int \Lambda\left(\frac{x}{t}\right)|t|\,\frac{k}{2}R(k(t-1))\,dt = \frac{k}{2}\int_{1-\frac{1}{k}}^{1+\frac{1}{k}} \Lambda\left(\frac{x}{t}\right)tdt.$$

The corresponding weight functions are derived in Corollary 110.
We proceed by considering x in three separate sets:

Case 1 $0 \leq x < 1 - \frac{1}{k}$.

$$
\begin{aligned}
G_k(x) &= \frac{k}{2} \int_{1-\frac{1}{k}}^{1+\frac{1}{k}} \Lambda\left(\frac{x}{t}\right) t dt = \frac{k}{2} \int_{1-\frac{1}{k}}^{1+\frac{1}{k}} \left(1 - \frac{x}{t}\right) t dt = \frac{k}{2k} \int_{1-\frac{1}{k}}^{1+\frac{1}{k}} (t - x)\, dt \\
&= \frac{k}{2} \left[\frac{1}{2}t^2 - tx\right]_{1-\frac{1}{k}}^{1+\frac{1}{k}} \\
&= \frac{k}{2} \left(\frac{1}{2}\left(1 + \frac{1}{k}\right)^2 - \left(1 + \frac{1}{k}\right) x\right) - \frac{1}{k}\left(-\left(1 - \frac{1}{k}\right) x\right) \\
&= \frac{k}{2} \left(\frac{1}{2}\left(1 + \frac{1}{k}\right)^2 - \left(1 + \frac{1}{k}\right) x - \frac{1}{2}\left(1 - \frac{1}{k}\right)^2 + \left(1 - \frac{1}{k}\right) x\right) \\
&= k\left(\frac{1}{k} - \frac{x}{k}\right) = (1 - x) = \Lambda(x).
\end{aligned}
$$

Case 2 $1 - \frac{1}{k} \leq x < 1 + \frac{1}{k}$. *Then*

$$
\begin{aligned}
G_k(x) = \frac{k}{2} \int_{x}^{1+\frac{1}{k}} \Lambda\left(\frac{x}{t}\right) t dt &= \frac{k}{2} \int_{x}^{1+\frac{1}{k}} \left(1 - \frac{x}{t}\right) t dt = \frac{k}{2} \int_{x}^{1+\frac{1}{k}} (t - x)\, dt \\
&= \frac{k}{2} \left[\frac{1}{2}t^2 - tx\right]_{x}^{1+\frac{1}{k}} \\
&= \frac{k}{2} \left(\left(\frac{1}{2}\left(1 + \frac{1}{k}\right)^2 - \left(1 + \frac{1}{k}\right) x\right) - \left(\frac{x^2}{2} - x^2\right)\right) \\
&= \frac{k}{2} \left(\frac{1}{2}\left(1 + \frac{1}{k}\right)^2 - \left(1 + \frac{1}{k}\right) x + \frac{1}{2}x^2\right) \\
&= \frac{k}{4} \left(\left(1 + \frac{1}{k}\right)^2 - 2\left(1 + \frac{1}{k}\right) x + x^2\right) \\
&= \frac{k}{4} \left(1 + \frac{1}{k} - x\right)^2.
\end{aligned}
$$

Case 3 $1 + \frac{1}{k} \leq x$. *Clearly* $G_k(x) = 0$.
From the three cases we conclude that

$$
G_k(x) = \begin{cases} \Lambda(x), & 0 \leq |x| \leq 1 - \frac{1}{k}, \\ \frac{k}{4}\left(1 + \frac{1}{k} - |x|\right)^2, & 1 - \frac{1}{k} \leq |x| \leq 1 + \frac{1}{k}, \\ 0, & 1 + \frac{1}{k} \leq |x|, \end{cases} \tag{3.36}
$$

and $G_k \in C\left(\mathbb{R}^1\right) \cap C^{(1)}\left(\mathbb{R}^1 \backslash 0\right) \cap PWC^{\infty}\left(\mathbb{R}^1 \backslash \{-1, 0, 1\}\right)$. *We can also write*

$$
G_k(x) = \Lambda(x) + \Delta_k(x),
$$

where

$$
\begin{aligned}
\Delta_k(x) &= \begin{cases} 0, & 0 \leq |x| \leq 1 - \frac{1}{k}, \\ \frac{k}{4}\left(1 - |x| + \frac{1}{k}\right)^2 - (1 - |x|), & 1 - \frac{1}{k} \leq |x| \leq 1, \\ \frac{k}{4}\left(1 - |x| + \frac{1}{k}\right)^2, & 1 \leq |x| \leq 1 + \frac{1}{k}, \\ 0, & 1 + \frac{1}{k} \leq |x|, \end{cases} \\
&= \begin{cases} 0, & 0 \leq |x| \leq 1 - \frac{1}{k}, \\ \frac{k}{4}\left(|x| - \left(1 - \frac{1}{k}\right)\right)^2, & 1 - \frac{1}{k} \leq |x| \leq 1, \\ \frac{k}{4}\left(1 + \frac{1}{k} - |x|\right)^2, & 1 \leq |x| \leq 1 + \frac{1}{k}, \\ 0, & 1 + \frac{1}{k} \leq |x|. \end{cases}
\end{aligned} \tag{3.37}
$$

We note that

$$\operatorname{supp}\Delta_k = \left[1-\frac{1}{k}, 1+\frac{1}{k}\right]; \quad 0 \le |\Delta_k(x)| \le \frac{1}{4k}.$$

3.3.2 Lipschitz continuity

The next two theorems will show that the continuous derivatives of the basis function are uniformly Lipschitz continuous of order 1 on $\mathbb{R}^1$. We start with the univariate case.

Theorem 118 *Let $w(\xi) = \frac{\xi^{2n}}{\Delta^{2l}\widehat{q}(\xi)}$ be the univariate central difference weight function introduced in Definition 97 and which satisfies property W2. If G_1 is the basis function of order zero generated by w the derivatives $D^k G_1(x)$, $k \le 2n-2$, are uniformly Lipschitz continuous of order 1 on $\mathbb{R}^1$. In fact*

$$\left|D^k G_1(x) - D^k G_1(y)\right| \le \left\|D^{k+1} G_\Lambda\right\|_\infty \left\|t^{2n-2-k} q\right\|_1 |x-y|, \quad x, y \in \mathbb{R}^1,$$

where G_Λ is given by 3.27.

Proof. From 3.29 of Theorem 111

$$\begin{aligned}\left|D^k G_1(x) - D^k G_1(y)\right| &\le \int\limits_{\mathbb{R}^1} \left|\left(D^k G_\Lambda\right)\left(\frac{x}{t}\right) - \left(D^k G_\Lambda\right)\left(\frac{y}{t}\right)\right| |t|^{2n-k-1} q(t)\, dt \\ &\le \int\limits_{\mathbb{R}^1} \left|\left(D^k G_\Lambda\right)\left(\frac{x}{t}\right) - \left(D^k G_\Lambda\right)\left(\frac{y}{t}\right)\right| |t|^{2n-k-1} q(t)\, dt.\end{aligned}$$

Since G_Λ is a constant scalar multiple of the extended B-spline, from Theorem 118 we have for dimension 1 and $k \le 2n-2$,

$$\left|D^k G_\Lambda(s) - D^k G_\Lambda(s')\right| \le \left\|D^{k+1} G\right\|_\infty |s-s'|, \quad s, s' \in \mathbb{R}^1,$$

and thus

$$\begin{aligned}\left|D^k G_1(x) - D^k G_1(y)\right| &\le \left\|D^{k+1} G_\Lambda\right\|_\infty \int\limits_{\mathbb{R}^1} \left|\frac{x}{t} - \frac{y}{t}\right| |t|^{2n-k-1} q(t)\, dt \\ &\le \left\|D^{k+1} G_\Lambda\right\|_\infty \left\|t^{2n-2-k} q\right\|_1 |x-y|.\end{aligned}$$

■

To deal with the multivariate case we will require a lemma concerning a distribution Taylor series expansion and which follows directly from Lemma 42 Chapter 1:

Lemma 119 *Suppose that $G \in C_B^{(0)}\left(\mathbb{R}^d\right)$ and the (distributional) derivatives $\{D^\alpha G\}_{|\alpha|=1}$ are bounded functions. Then*

$$G(x) - G(y) = \sum_{|\alpha|=1} (x-y)^\alpha \int_0^1 (D^\alpha G)(x - t(x-y))\, dt,$$

and

$$|G(x) - G(y)| \le \sqrt{d} \max_{|\alpha|=1} \|D^\alpha G\|_\infty |x-y|. \tag{3.38}$$

Now we can show that the continuous derivatives of the tensor product central difference basis functions are also Lipschitz continuous of order 1.

Theorem 120 *Let $G(x) = \prod\limits_{k=1}^{d} G_1(x_k)$ be a central difference tensor product basis function with parameters n, l, $q(\cdot)$, and q is bounded when $n=1$. Then $G_1 \in C_B^{(2n-2)}$, $D^{2n-1}G_1$ is bounded and we have the estimates*

$$|G(x) - G(y)| \le \sqrt{d} G_1(0)^{d-1} |DG_1|_\infty |x-y|, \quad x, y \in \mathbb{R}^d, \tag{3.39}$$

and if $0 < |\beta| \le 2n-2$

$$\left|D^\beta G(x) - D^\beta G(y)\right| \le \sqrt{d} \left(\max_{k=1}^{d} \left|D_k D^\beta G\right|_\infty\right) |x-y|, \quad x, y \in \mathbb{R}^d. \tag{3.40}$$

Proof. By Theorem 111, $G_1 \in C_B^{(2n-2)}$ and by Theorem 115, $D^{2n-1}G_1$ is bounded.

Thus $G \in C_B^{(0)}\left(\mathbb{R}^d\right)$ and the derivatives $\{D^\alpha G\}_{|\alpha|=1}$ are bounded functions. Thus G satisfies the conditions of Lemma 119 and so the estimate of that lemma holds i.e. for all $x, y \in \mathbb{R}^d$

$$|G(x) - G(y)| \leq \sqrt{d}\left(\max_{|\alpha|=1} |D^\alpha G|_\infty\right)|x-y| \leq \sqrt{d}\,|G_1|_\infty^{d-1}|DG_1|_\infty |x-y|.$$

Finally, from 1.30, $|G_1(x)| \leq G_1(0)$ which proves 3.39.

To prove 3.40 observe that if $0 < k \leq 2n-2$ then $D^k G_1 \in C_B^{(2n-2-k)}\left(\mathbb{R}^d\right)$ and $D^{k+1}G_1$ is bounded. Thus if $0 < |\beta| \leq 2n-2$ then $D^\beta G \in C_B^{(0)}\left(\mathbb{R}^d\right)$ and the derivatives $\left\{D^\alpha D^\beta G\right\}_{|\alpha|=1}$ are bounded functions. Thus $D^\beta G$ satisfies the conditions of Lemma 119 and so the estimate of that lemma holds i.e. for all $x, y \in \mathbb{R}^d$

$$\left|D^\beta G(x) - D^\beta G(y)\right| \leq \sqrt{d}\left(\max_{|\alpha|=1}\left|D^\alpha D^\beta G\right|_\infty\right)|x-y| \leq \sqrt{d}\left(\max_{k=1}^{d}\left|D_k D^\beta G\right|_\infty\right)|x-y|.$$

■

3.4 Weight and basis function convergence results

In the preamble to the definition of a central difference weight function in Subsection 3.2.1 several assertions were made regarding the convergence of sequences of central difference weight and basis functions to extended B-spline weight and basis functions. This section is devoted to proving these assertions.

3.4.1 Pointwise basis function convergence

The next lemma shows that the tensor product extended B-spline basis functions are Lipschitz continuous of order 1.

Lemma 121 *(Theorem 43 Chapter 1) Let $G(x) = \prod_{k=1}^{d} G_1(x_k)$ be a tensor product* ***extended B-spline*** *basis function, as described in Theorem 29. We then have the estimate*

$$|G(x) - G(y)| \leq \sqrt{d}G_1(0)^{d-1}\|DG_1\|_\infty |x-y|, \quad x, y \in \mathbb{R}^d.$$

Theorem 122 *Suppose $l, n \geq 1$ are integers which satisfy $1 \leq n \leq l$. Let $\psi \in L^1\left(\mathbb{R}^1\right)$ be a function which satisfies $\psi(x) \geq 0$, $\psi(0) = 1$ and*

$$\int_{\mathbb{R}^1} |t|^{2n}\psi(t)\,dt < \infty. \tag{3.41}$$

Then the sequence of functions $q_k(t) = \sqrt{2\pi}2^{-2l}k\psi(k(t-2))$, $\quad k = 1, 2, 3, \ldots$ generates a sequence of central difference basis functions, say G_k, which converges uniformly pointwise to the extended B-spline basis function G_s with parameters l and n. In fact, there exists a constant $C = C(G_s)$ which depends only on G_s such that

$$|G_k(x) - G_s(x)| \leq \frac{2^{2n-1}C(G_s)}{k}, \quad x \in \mathbb{R}^1.$$

Proof. It is clear that functions q_k satisfy the conditions of Theorem 98 so that each function $\frac{\xi^{2n}}{\Delta^{2l}\widehat{q_k}(\xi)}$ is a central difference weight function with parameters l and n. By Theorem 111 the corresponding basis functions are

$$G_k(x) = \int G_s\left(\frac{x}{t}\right)|t|^{2n-1}k\psi(k(t-2))\,dt,$$

where $G_s = \frac{(-1)^{(l-n)}}{\sqrt{2\pi}} 2^{-2l} D^{2(l-n)} (*\Lambda)^l$. Since $\int \psi = 1$ it follows that

$$\begin{aligned}
G_k(x) - G_s(x) &= \int_{1-\frac{1}{k}}^{1+\frac{1}{k}} \left(G_s\left(\frac{x}{t}\right) |t|^{2n-1} - G_s(x) \right) k\psi(k(t-2))\, dt \\
&= \int_{1-\frac{1}{k}}^{1+\frac{1}{k}} \left(G_s\left(\frac{x}{t}\right) |t|^{2n-1} - G_s(x) |t|^{2n-1} \right) k\psi(k(t-2))\, dt + \\
&\quad + \int_{1-\frac{1}{k}}^{1+\frac{1}{k}} \left(G_s(x) |t|^{2n-1} - G_s(x) \right) k\psi(k(t-2))\, dt \\
&= \int_{1-\frac{1}{k}}^{1+\frac{1}{k}} \left(G_s\left(\frac{x}{t}\right) - G_s(x) \right) |t|^{2n-1} k\psi(k(t-2))\, dt + \\
&\quad + \int_{1-\frac{1}{k}}^{1+\frac{1}{k}} G_s(x) \left(|t|^{2n-1} - 1 \right) k\psi(k(t-2))\, dt,
\end{aligned}$$

so that by the Lipschitz continuity estimate on $\mathbb{R}^1$ derived for the extended B-spline in Lemma 121

$$\begin{aligned}
|G_k(x) - G_s(x)| &\le \int_{1-\frac{1}{k}}^{1+\frac{1}{k}} \left| G_s\left(\frac{x}{t}\right) - G_s(x) \right| t^{2n-1} k\psi(k(t-2))\, dt + \\
&\quad + \int_{1-\frac{1}{k}}^{1+\frac{1}{k}} |G_s(x)| \left| t^{2n-1} - 1 \right| k\psi(k(t-2))\, dt \\
&\le c \int_{1-\frac{1}{k}}^{1+\frac{1}{k}} \left| \frac{x}{t} - x \right| t^{2n-1} k\psi(k(t-2))\, dt + \\
&\quad + \|G_s\|_\infty \int_{1-\frac{1}{k}}^{1+\frac{1}{k}} \left| t^{2n-1} - 1 \right| k\psi(k(t-2))\, dt \\
&= c|x| \int_{1-\frac{1}{k}}^{1+\frac{1}{k}} |1-t|\, t^{2n-2} k\psi(k(t-2))\, dt + \\
&\quad + \|G_s\|_\infty \int_{1-\frac{1}{k}}^{1+\frac{1}{k}} |t-1| \left(1 + t + \ldots + t^{2n-2}\right) k\psi(k(t-2))\, dt \\
&\le \frac{c 2^{2n-1} |x|}{k} + \frac{\|G_s\|_\infty 2^{2n-1}}{k},
\end{aligned}$$

and since G_s has compact support we have uniform convergence. ■

3.4.2 Weight function convergence

Recall that at the end of the Subsection 3.2.1 it was suggested that if $\psi \in L^1$ satisfies $\psi(x) \ge 0$ and $\int \psi = 1$ and an extra condition then the sequence of functions $q_k(t) = \frac{\sqrt{2\pi}}{2^{2l}} k\psi(k(t-2)) \in L^1$ is such that $\frac{\Delta^{2l} \widehat{q_k}(\xi)}{\xi^{2n}}$ converges to $\frac{\sin^{2l}\xi}{\xi^{2n}}$ in the sense of tempered distributions. Here $\frac{\sin^{2l}\xi}{\xi^{2n}} = \frac{1}{w_s(\xi)} \in L^1$ where w_s is the extended B-spline weight function with parameters n, l. The extra condition referred to is 3.41 i.e. $\int_{|t| \ge R} |t|^{2n} q_k(t)\, dt < \infty$ for some $R \ge 0$, which is equivalent to $\int |t|^{2n} \psi(t)\, dt < \infty$ since $\int \psi < \infty$. This assumption is stronger than 3.3 which was the used in Theorem 98 to ensure that a central difference weight function has property W2. We will prove L^1 convergence and this implies convergence in S'.

Assume $\xi > 0$. Then using the change of variables: $s = \xi t/2$, $dt = \frac{2}{\xi} ds$

$$\begin{aligned}
\frac{1}{w_k(\xi)} &= \frac{2^{2(l-n)}}{\sqrt{2\pi}} \int \frac{t^{2n} q_k(t)}{w_s(\xi t/2)} dt \\
&= \int \frac{(t/2)^{2n} k\psi(k(t-2))}{w_s(\xi t/2)} dt \\
&= \int \left(\frac{s}{\xi}\right)^{2n} \frac{\frac{2k}{\xi}\psi\left(\frac{2k}{\xi}(s-\xi)\right)}{w_s(s)} ds \\
&= \frac{1}{\xi^{2n}} \int \frac{2k}{\xi}\psi\left(\frac{2k}{\xi}(s-\xi)\right) \sin^{2l} s \; ds \\
&= \frac{1}{\xi^{2n}} \int \frac{2k}{\xi}\psi\left(\frac{2k}{\xi}(s-\xi)\right) \left(\sin^{2l} s - \sin^{2l} \xi\right) ds + \\
&\qquad + \frac{1}{\xi^{2n}} \int \frac{2k}{\xi}\psi\left(\frac{2k}{\xi}(s-\xi)\right) \sin^{2l} \xi \; ds \\
&= \frac{1}{\xi^{2n}} \int \frac{2k}{\xi}\psi\left(\frac{2k}{\xi}(s-\xi)\right) \left(\sin^{2l} s - \sin^{2l} \xi\right) ds + \frac{\sin^{2l} \xi}{\xi^{2n}} \int \psi \\
&= \frac{1}{\xi^{2n}} \int \frac{2k}{\xi}\psi\left(\frac{2k}{\xi}(s-\xi)\right) \left(\sin^{2l} s - \sin^{2l} \xi\right) ds + \frac{\sin^{2l} \xi}{\xi^{2n}},
\end{aligned}$$

so that

$$\frac{1}{w_k(\xi)} - \frac{1}{w_s(\xi)} = \frac{1}{\xi^{2n}} \int \frac{2k}{\xi}\psi\left(\frac{2k}{\xi}t\right) \left(\sin^{2l}(t+\xi) - \sin^{2l} \xi\right) dt.$$

Now split the domain of integration about $|\xi| = R$: since $\int \psi = 1$,

$$\begin{aligned}
\int\limits_{|\cdot| \geq R} \left| \frac{1}{w_k} - \frac{1}{w_s} \right| &\leq \int\limits_{|\xi| \geq R} \frac{1}{\xi^{2n}} \int \frac{2k}{\xi}\psi\left(\frac{2k}{\xi}t\right) \left|\sin^{2l}(t+\xi) - \sin^{2l} \xi\right| dt d\xi \\
&\leq 2 \int\limits_{|\xi| \geq R} \frac{1}{\xi^{2n}} \int \frac{2k}{\xi}\psi\left(\frac{2k}{\xi}t\right) dt d\xi \\
&\leq 2 \int\limits_{|\xi| \geq R} \frac{d\xi}{\xi^{2n}} \\
&= 4 \int\limits_{\xi \geq R} \frac{d\xi}{\xi^{2n}} \\
&= \frac{4}{2n-1} \frac{1}{R^{2n-1}}.
\end{aligned}$$

If G_s is the B-spline basis function we can write

$$\begin{aligned}
\frac{1}{w_k(\xi)} - \frac{1}{w_s(\xi)} &= \frac{1}{\xi^{2n}} \int \frac{2k}{\xi}\psi\left(\frac{2k}{\xi}(s-\xi)\right)\left(\sin^{2l} s - \sin^{2l}\xi\right) ds \\
&= \frac{1}{\xi^{2n}} \int \frac{2k}{\xi}\psi\left(\frac{2k}{\xi}t\right)\left(\sin^{2l}(\xi+t) - \sin^{2l}\xi\right) dt \\
&= \int \frac{2k}{\xi}\psi\left(\frac{2k}{\xi}t\right)\left(\frac{\sin^{2l}(\xi+t)}{\xi^{2n}} - \frac{\sin^{2l}\xi}{\xi^{2n}}\right) dt \\
&= \int \frac{2k}{\xi}\psi\left(\frac{2k}{\xi}t\right)\left(\frac{(t+\xi)^{2n}}{\xi^{2n}}\widehat{G_s}(\xi+t) - \widehat{G_s}(\xi)\right) dt \\
&= \int \frac{2k}{\xi}\psi\left(\frac{2k}{\xi}t\right)\left(\left(1+\frac{t}{\xi}\right)^{2n}\widehat{G_s}(\xi+t) - \widehat{G_s}(\xi)\right) dt \\
&= \int \frac{2k}{\xi}\psi\left(\frac{2k}{\xi}t\right)\left(\sum_{m\le 2n}\binom{2n}{k}\left(\frac{t}{\xi}\right)^m\widehat{G_s}(\xi+t) - \widehat{G_s}(\xi)\right) dt \\
&= \sum_{2\le m\le 2n}\binom{2n}{m}\int \frac{2k}{\xi}\left(\frac{t}{\xi}\right)^m\psi\left(\frac{2k}{\xi}t\right)\widehat{G_s}(\xi+t)\,dt + \\
&\quad + \int \frac{2k}{\xi}\psi\left(\frac{2k}{\xi}t\right)\left(\widehat{G_s}(\xi+t) - \widehat{G_s}(\xi)\right) dt \\
&= \sum_{2\le m\le 2n}\binom{2n}{m}\int \frac{2k}{\xi}\left(\frac{t}{\xi}\right)^m\psi\left(\frac{2k}{\xi}t\right)\widehat{G_s}(\xi+t)\,dt + \\
&\quad + \int \frac{2k}{\xi}\psi\left(\frac{2k}{\xi}t\right)\left(\widehat{G_s}(\xi+t) - \widehat{G_s}(\xi)\right) dt. \qquad (3.42)
\end{aligned}$$

$$\begin{aligned}
\int\left|\frac{1}{w_k} - \frac{1}{w_s}\right| \le \sum_{2\le m\le 2n}\binom{2n}{m}\int_{|\xi|\le R}\int\left|\frac{2k}{\xi}\left(\frac{t}{\xi}\right)^m\right|\psi\left(\frac{2k}{\xi}t\right)\widehat{G_s}(\xi+t)\,dtd\xi + \\
+\int_{|\xi|\le R}\int\frac{2k}{|\xi|}\psi\left(\frac{2k}{\xi}t\right)\left|\widehat{G_s}(\xi+t) - \widehat{G_s}(\xi)\right| dtd\xi + \frac{4}{2n-1}\frac{1}{R^{2n-1}}.
\end{aligned}$$

We will consider the last integral in 3.42 first. Since

$$\begin{aligned}
\left|\widehat{G_s}(\xi+t) - \widehat{G_s}(\xi)\right| = \left|\widehat{e^{i(t,\cdot)}G_s}(\xi) - \widehat{G_s}(\xi)\right| &= \left|\left(\left(e^{i(t,\cdot)}-1\right)G_s\right)^\wedge(\xi)\right| \\
&= |t|\ \left|\left(\left(\frac{e^{i(t,\cdot)}-1}{t}\right)G_s\right)^\wedge(\xi)\right| \\
&\le |t|\ \|xG_s(x)\|_1,
\end{aligned}$$

it follows that

$$\begin{aligned}
\int \frac{2k}{|\xi|}\psi\left(\frac{2k}{\xi}t\right)\left|\widehat{G_s}(\xi+t) - \widehat{G_s}(\xi)\right| dt &\le \|xG_s(x)\|_1 \int \frac{2k}{|\xi|}|t|\,\psi\left(\frac{2k}{\xi}t\right) dt \\
&= \frac{|\xi|}{2k}\|xG_s(x)\|_1\|s\psi(s)\|_1,
\end{aligned}$$

which exists since $\psi \in L^1$ and $\int_{|s|\ge R}|s|^{2n}\psi(s)\,ds < \infty$. Thus

$$\begin{aligned}
\int_{|\xi|\le R}\int\frac{2k}{|\xi|}\psi\left(\frac{2k}{\xi}t\right)\left|\widehat{G_s}(\xi+t) - \widehat{G_s}(\xi)\right| dtd\xi &\le \int_{|\xi|\le R}\frac{|\xi|}{2k}\|xG_s(x)\|_1\|s\psi(s)\|_1\,d\xi \\
&= \frac{R^2}{2k}\|xG_s(x)\|_1\|s\psi(s)\|_1. \qquad (3.43)
\end{aligned}$$

Regarding the integrals under the summation sign of 3.42, for $m = 2, 3, \ldots, 2n$ we make the change of variable $s = \frac{2k}{\xi} t$ to get

$$\begin{aligned}\left| \int_0^R \int \frac{2k}{\xi} \left(\frac{t}{\xi} \right)^m \psi \left(\frac{2k}{\xi} t \right) \widehat{G_s} (\xi + t) \, dt d\xi \right| &< \int_0^R \int \frac{2k}{\xi} \left(\frac{|t|}{\xi} \right)^m \psi \left(\frac{2k}{\xi} t \right) dt d\xi \\ &= \frac{1}{(2k)^m} \int_0^R \int |s|^m \psi (s) \, ds d\xi \\ &= \frac{R}{(2k)^m} \left\| s^m \psi (s) \right\|_1 ,\end{aligned}$$

and similarly

$$\begin{aligned}\left| \int_{-R}^0 \int \frac{2k}{\xi} \left(\frac{t}{\xi} \right)^m \psi \left(\frac{2k}{\xi} t \right) \widehat{G_s} (\xi + t) \, dt d\xi \right| &< \frac{R}{(2k)^m} \int |s|^m \psi (s) \, ds \\ &= \frac{R}{(2k)^m} \left\| s^m \psi (s) \right\|_1 ,\end{aligned}$$

so that

$$\begin{aligned}\left| \int_{|\xi| < R} \int \frac{2k}{\xi} \left(\frac{t}{\xi} \right)^m \psi \left(\frac{2k}{\xi} t \right) \widehat{G_s} (\xi + t) \, dt d\xi \right| &< \frac{2R}{(2k)^m} \int |s|^m \psi (s) \, ds \\ &= \frac{2R}{(2k)^m} \left\| s^m \psi (s) \right\|_1 . \end{aligned} \tag{3.44}$$

Inequalities 3.43 and 3.44 can now be used to estimate the right side of 3.42: if $R > 2$

$$\begin{aligned}\int \left| \frac{1}{w_k} - \frac{1}{w_s} \right| &< \sum_{2 \le m \le 2n} \binom{2n}{m} \frac{2R}{(2k)^m} \left\| s^m \psi (s) \right\|_1 + \frac{R^2}{2k} \left\| x G_s (x) \right\|_1 \left\| s \psi (s) \right\|_1 + \\ &\quad + \frac{4}{2n-1} \frac{1}{R^{2n-1}} \\ &= \frac{R^2}{2k} \left(\sum_{2 \le m \le 2n} \binom{2n}{m} + \left\| x G_s (x) \right\|_1 \right) \max_{m=1}^{2n} \left\{ \left\| s^m \psi (s) \right\|_1 \right\} + \frac{4}{R^{2n-1}} \\ &= \frac{R^2}{2k} \left(2^{2n} - 1 + \left\| x G_s (x) \right\|_1 \right) \max_{m=1}^{2n} \left\{ \left\| s^m \psi (s) \right\|_1 \right\} + \frac{4}{R^{2n-1}},\end{aligned}$$

and it is clear that given $\varepsilon > 0$ the right side can be made less than ε by first choosing R such that $\frac{4}{R^{2n-1}} < \frac{\varepsilon}{2}$ and then choosing k so that the remaining terms also do not exceed $\varepsilon/2$. Thus we have shown that $\frac{1}{w_k} \to \frac{1}{w_s}$ in L^1 and hence in S'.

3.5 Local data functions

In this section we will use the generalized local data function results of Theorem 91 and Corollary 92 of Chapter 2 to characterize locally the data functions for the tensor product central difference basis functions as Sobolev-like spaces. These results supply important information about the data functions and makes it easy to choose data functions for numerical experiments concerning the zero order basis function interpolation and smoothing problems discussed later in Chapters 2, 4 and 5.

3.5.1 Tensor product weight functions

Here we are interested in tensor product weight functions e.g. the multivariate central difference weight functions and the extended B-splines 1.20. We want a result which allows us to prove that there exists integral n such that the local Sobolev-like space

$$H_s^n (\Omega) = \left\{ u \in L^2 (\Omega) : D^\alpha u \in L^2 (\Omega) \;\; for \; \alpha \le n \right\},$$

of Definition 81 is a subset of the restriction space $X_w^0(\Omega)$ (recall Definition 86). Further we want this result to (essentially) involve **only proving estimates about the univariate weight function**. This result is Theorem 123 and will allow us to justify using $H_s^n(\Omega)$ functions as data functions in numerical experiments.

Further, if w_1 satisfies the assumptions of Corollary 124 then $H_s^n(\Omega) = X_w^0(\Omega)$ as sets and their norms are equivalent.

Theorem 123 *Let w be a tensor product weight function on $\mathbb{R}^d$ generated by the univariate weight function w_1. Suppose:*

1. *$\Omega \subset \mathbb{R}^d$ is an open, bounded set which permits one of the extension operators $\mathcal{E}_\Omega : H_s^n(\Omega) \to H_s^n$ described in parts 5 to 7 of Remark 82.*

2. *There exists a bounded function $g_1 : \mathbb{R}^1 \to \mathbb{C}^1$ and constants $C_k^{(1)} \geq 0$ such that*
$$|g_1(t)|^2 w_1(t) \leq \sum_{m \leq n} C_m^{(1)} t^{2m}, \quad t \in \mathbb{R}^1, \tag{3.45}$$
and if we define $g : \mathbb{R}^d \to \mathbb{C}^1$ by $g(\xi) = g_1(\xi_1) \times \ldots \times g_1(\xi_d)$ then
$$\operatorname{supp}(g(D)u - u) \subset \left([-1,1]^d\right)^c, \quad u \in H_0^n\left((-1,1)^d\right), \tag{3.46}$$
where $g(D)u = (g\widehat{u})^\vee$.

3. *There exists an open cube C (with center c) containing $\overline{\Omega}$ and an integer $m > 0$ for which*
$$w_1(t/m) \leq c_m^{(1)} w_1(t), \qquad t \in \mathbb{R}^1, \tag{3.47}$$
for some constant $c_m^{(1)} > 0$ and $(C - c)/m \subset [-1,1]^d$.

Then $H_s^n(\Omega) \hookrightarrow X_w^0(\Omega)$ and this inclusion is also valid for any scaled weight function with appropriately scaled Ω.

Proof. The proof involves showing the assumptions required by Theorem 91 are met. Firstly, observe that assumption 1 of this theorem is assumption 1 of Theorem 91. Next define the bounded function $g : \mathbb{R}^d \to \mathbb{C}^1$ by $g(\xi) = \prod_{k=1}^{d} g(\xi_k)$. Then if $C_\alpha = C_{\alpha_1}^{(1)} C_{\alpha_2}^{(1)} \ldots C_{\alpha_d}^{(1)}$,
$$|g(\xi)|^2 w(\xi) = \prod_{k=1}^{d} |g_1(\xi_k)|^2 w_1(\xi_k) \leq \prod_{k=1}^{d} \left(\sum_{\alpha_k \leq n} C_{\alpha_k}^{(1)} \xi_k^{2\alpha_k} \right) = \sum_{\alpha \leq n} C_\alpha^{(1)} \xi^{2\alpha} \leq \sum_{\alpha \leq n} C_\alpha \xi^{2\alpha},$$
which confirms 2.75 and as a consequence 3.46 implies that assumption 2 is also satisfied.

Finally, if $c_m = \left(c_m^{(1)}\right)^d$ and inequality 3.47 is valid then
$$w(\xi/m) = \prod_{k=1}^{d} w_1(\xi_k/m) \leq \prod_{k=1}^{d} c_m^{(1)} w_1(\xi_k) = c_m w(\xi),$$
which means that assumption 3 of Theorem 91 is also satisfied. The conclusion of this theorem now follows from that of Theorem 91. ■

This corollary supplies a condition under which $X_w^0(\Omega) = H_s^n(\Omega)$ as sets with equivalent norms.

Corollary 124 *Suppose the assumptions of Theorem 123 hold. Also suppose $\Omega \subset R_\Omega$ where R_Ω is an open rectangle and assume that there exist constants $c_k > 0$ such that the univariate weight function w_1 satisfies*
$$w_1(t) \geq c_k t^{2k}, \quad t \in \bigcup_{j=1}^{d} \pi_j(R_\Omega), \; k = 0, 1, 2, \ldots, n, \tag{3.48}$$
where $\pi_j(x) = x_j$ for $x = (x_j) \in \mathbb{R}^d$.

Then $X_w^0(\Omega) = H_s^n(\Omega)$ as sets and their norms are equivalent. This result is also valid for any scaled weight function with appropriately scaled Ω.

Proof. Suppose $\xi \in \Omega$ and $\alpha \leq n$ where α is a multi-index. Set $c_\alpha = \prod\limits_{j=1}^{d} c_{\alpha_j}$. Then $\xi \in R_\Omega$ and

$$w(\xi) = \prod_{m=1}^{d} w_1(\xi_m) \geq \prod_{j=1}^{d} c_{\alpha_j} \xi_j^{2\alpha_j} = c_\alpha \xi^{2\alpha},$$

and this corollary now follows from Corollary 92. ■

Examples

We present two applications of the theory of data functions introduced in the last subsection to tensor product central difference weight functions.

In the first example the strong integral condition 3.11 is imposed on the weight function because this condition is associated with simple upper and lower bounds on w which imply $X_w^0(\Omega) = H_s^n(\Omega)$ as sets.

Theorem 125 *Suppose w is a tensor product central difference weight function on $\mathbb{R}^d$ with parameters n, l, q which satisfies the condition 3.11 i.e. $\int_{|t| \geq R} t^{2l} q(t)\, dt < \infty$. Further suppose that Ω is an open, bounded set which permits one of the extension operators $\mathcal{E}_\Omega : H_s^n(\Omega) \to H_s^n$ described in parts 5 to 7 of Remark 82.*

Then $X_w^0(\Omega) = H_s^n(\Omega)$ as sets and the norms are equivalent.

Proof. Denote by w_1 the univariate weight function. From Corollary 109, given any $a > 0$ there exist constants $c_a, c_a', k_a, k_a' > 0$ such that

$$k_a t^{2n} \leq w_1(t) \leq k_a' t^{2n}, \quad |t| \geq a, \tag{3.49}$$

and

$$\frac{c_a}{t^{2(l-n)}} \leq w_1(t) \leq \frac{c_a'}{t^{2(l-n)}}, \quad |t| \leq a. \tag{3.50}$$

We will verify that 3.49 and 3.50 imply that the conditions of Theorem 123 and Corollary 124 hold.

Noting Remark 94 above, if we choose $g(\xi) = (\sin \xi)^{l-n}$ then 3.46 is satisfied. In fact, if we (arbitrarily) choose $a = 1$ then: $|t| \leq 1$ implies $|g_1(t)|^2 w_1(t) \leq \frac{(\sin t)^{2(l-n)}}{t^{2(l-n)}} c_1' \leq c_1'$and $|t| \geq 1$ implies $|g_1(t)|^2 w_1(t) \leq (\sin t)^{2(l-n)} k_1' t^{2n} \leq k_1' t^{2n}$, so that 3.45 holds.

Now we show that 3.47 holds for all integer $m \geq 1$: if $|t| \geq 1$ then from 3.49 $w_1(t/m) \leq \frac{k_1'}{m^{2n}} t^{2n}$ and $1/w_1(t) \leq \frac{1}{k_1 t^{2n}}$ so that $\frac{w_1(t/m)}{w_1(t)} \leq \frac{k_1'}{m^{2n} k_1}$; if $|t| < 1$ then from 3.50, $w_1(t/m) \leq \frac{m^{2(l-n)} c_1'}{t^{2(l-n)}}$ and $1/w_1(t) \leq \frac{1}{c_1} t^{2(l-n)}$ and thus $\frac{w_1(t/m)}{w_1(t)} \leq \frac{m^{2(l-n)} c_1'}{c_1}$. Hence 3.47 holds for all integer $m \geq 1$.

The next step is to show that 3.48 holds: here we choose $a = \max \left\{ |t| : t \in \bigcup\limits_{j=1}^{d} \pi_j\left(\overline{R_\Omega}\right) \right\}$ so that if $t \in \bigcup\limits_{j=1}^{d} \pi_j\left(\overline{R_\Omega}\right)$ and $\alpha \leq n$, from 3.50

$$w_1(t) \geq \frac{c_a}{t^{2(l-n)}} = \frac{c_a}{a^{2(l-n)}} \frac{1}{(t/a)^{2(l-n)}} \geq \frac{c_a}{a^{2(l-n)}} \geq \frac{c_a}{a^{2(l-n)}} \left(\frac{t}{a}\right)^{2\alpha} = \frac{c_a}{a^{2(\alpha+l-n)}} \xi^{2\alpha},$$

as required. ■

The last result assumed that $\int_{|t| \geq R} t^{2l} q(t)\, dt < \infty$. We now weaken this assumption to $\int_{|t| \geq R} t^{2n} q(t)\, dt < \infty$ and see what extra assumption is required to ensure that $X_w^0(\Omega) = H_s^l(\Omega)$.

Theorem 126 *Suppose w is a central difference weight function on $\mathbb{R}^d$ with parameters n, l, q which satisfies the condition 3.7 i.e. $\int_{|t| \geq R} t^{2n} q(t)\, dt < \infty$. Further suppose that Ω is an open, bounded set which permits one of the extension operators $\mathcal{E}_\Omega : H_s^n(\Omega) \to H_s^n$ described in parts 5 to 7 of Remark 82.*

Suppose, in addition, there exist constants $r, k_1 > 0$ such that

$$w(\xi) \geq \frac{k_1}{\xi^{2(l-n)}}, \quad |\xi| \leq r. \tag{3.51}$$

Then $X_w^0(\Omega) = H_s^n(\Omega)$ as sets and the norms are equivalent.

Proof. We first verify that 3.49 and 3.50 imply that the conditions of Theorem 91 and Corollary 92 hold for the case of one dimension. From 3.9 there exists a constant $C_1 > 0$ such that

$$w(\xi) \geq \frac{\xi^{2n}}{C_1}, \quad \xi \in \mathbb{R}^1, \tag{3.52}$$

and from 3.17 there exist constants $c_1, c_2 > 0$ and $0 \leq a < b$ such that

$$w(\xi) \leq \begin{cases} \frac{c_2}{\xi^{2(l-n)}}, & |\xi| \leq \frac{\pi}{b-a}, \\ 2c_1 \xi^{2n}, & |\xi| > \frac{\pi}{b-a}. \end{cases} \tag{3.53}$$

Using these two estimates and the additional assumption 3.51 we can derive inequalities of the form 3.49 and 3.50 and so prove this theorem in a similar manner to Theorem 125. ■

3.6 Interpolant convergence to its data function

We will use the results of 2 to derive orders for the pointwise converge of the minimal norm interpolant $\mathcal{I}_X f$ to its data function f. These estimates will have the form

$$|f(x) - (\mathcal{I}_X f)(x)| \leq k_G \|f\|_{w,0} (h_{X,\Omega})^m, \quad x \in \overline{\Omega},$$

where $h_{X,\Omega} = \sup_{\omega \in \Omega} \operatorname{dist}(\omega, X) < h_{\Omega,\kappa}$ measures the density (sparsity?) of the independent data X which is contained in the data region Ω. Here m is the order of convergence.

The results of Chapter 2 can be classified according to whether they require the independent data X to be unisolvent and a full discussion of unisolvency is incorporated into Subsection 2.5.3.

It turns out that all the convergence orders obtained here for the central difference weight functions are identical to those obtained for the extended B-splines.

3.6.1 Convergence estimates derived without assuming unisolvent data sets

Following the approach of Chapter 2 we consider two types of convergence results for the case when the data is not assumed to be unisolvent. These are imaginatively called Type 1 and Type 2.

Type 1 convergence estimates

Recall that Type 1 convergence estimates are those obtained using Theorem 59, where there are no *a priori* assumptions made about the weight function parameter κ but the smoothness condition 3.54 is applied to the basis function near the origin and this will allow a uniform order of convergence estimate to be obtained for the interpolant in a closed, bounded, infinite data region.

Theorem 127 *(Copy of Theorem 59) Suppose the weight function w has property W2 and that G is the basis function generated by w. Assume that for some $s > 0$ and constants $C_G, h_G > 0$ the basis function satisfies*

$$G(0) - \operatorname{Re} G(x) \leq C_G |x|^{2s}, \quad |x| < h_G. \tag{3.54}$$

Let $\mathcal{I}_X$ be the minimal norm interpolant mapping with the independent data set X contained in the closed, bounded, infinite set K, and let $k_G = (2\pi)^{-\frac{d}{4}} \sqrt{2C_G}$. Then for any data function $f \in X_w^0$ it follows that $\sqrt{(f - \mathcal{I}_X f, f)_{w,0}} \leq \|f\|_{w,0}$ and

$$|f(x) - \mathcal{I}_X f(x)| \leq k_G \sqrt{(f - \mathcal{I}_X f, f)_{w,0}} (h_{X,K})^s, \quad x \in K, \tag{3.55}$$

when $h_{X,K} = \max_{x \in K} \operatorname{dist}(x, X) < h_G$ i.e. the order of convergence is at least s.

In the next result we derive constants s, C_G and h_G for which the central difference basis function satisfies equation 3.54.

Theorem 128 *Suppose the weight function w is a tensor product central difference weight function with parameters n, l, $q(\cdot)$, and q is bounded if $n = 1$. Let G be the basis function generated by w.*

Then the basis function satisfies the estimate 3.54 for $s = 1/2$, $h_G = \infty$,

$$C_G = G_1(0)^{d-1} \|DG_1\|_\infty \sqrt{d},$$

and

$$k_G = (2\pi)^{-\frac{d}{4}} \sqrt{2C_G} = (2\pi)^{-\frac{d}{4}} \sqrt{2G_1(0)^{d-1} \|DG_1\|_\infty} \sqrt[4]{d}.$$

Proof. Let G_1 be the univariate basis function. By Theorem 120, $G_1 \in C_B^{(0)}$, DG_1 is bounded and

$$|G(x) - G(y)| \le \sqrt{d} G_1(0)^{d-1} \|DG_1\|_\infty |x - y|, \quad x, y \in \mathbb{R}^d,$$

so that we can choose $C_G = G_1(0)^{d-1} \|DG_1\|_\infty \sqrt{d}$ and

$$k_G = (2\pi)^{-\frac{d}{4}} \sqrt{2C_G} = (2\pi)^{-\frac{d}{4}} \sqrt{2G_1(0)^{d-1} \|DG_1\|_\infty} \sqrt[4]{d}.$$

∎

Type 2 convergence estimates

The Type 1 estimates of Theorem 127 considered the case where the weight function had property W2 for some $\kappa \ge 0$ and an extra condition was applied to the basis function. Type 2 convergence estimates only assume $\kappa \ge 1$. In the next theorem we only assume that $\kappa \ge 1$ and derive an order of convergence estimate of 1, as well as a doubled convergence estimate of 2 for the Riesz data functions R_y. Note that the constant k_G in the next Theorem is defined to match the constant k_G for the Type 1 convergence estimate 3.55.

Theorem 129 *(Copy of Theorem65) Suppose the weight function w has property W2 for a parameter $\kappa \ge 1$. Suppose also that $\mathcal{I}_X f$ is the minimal interpolant on X of the data function $f \in X_w^0$ and that K is a bounded, closed, infinite subset of $\mathbb{R}^d$, $X \subset K$ and $h_{X,K} = \max_{x \in K} \operatorname{dist}(x, X)$.*

Then

$$|f(x) - (\mathcal{I}_X f)(x)| \le k_G \sqrt{(f - \mathcal{I}_X f, f)_{w,0}}\, h_{X,K}, \quad x \in K,$$

where

$$k_G^2 = (2\pi)^{-d} \int \frac{|\cdot|^2}{w} = -(2\pi)^{-\frac{d}{2}} \left(\nabla^2 G\right)(0), \quad \sqrt{(f - \mathcal{I}_X f, f)_{w,0}} \le \|f\|_{w,0},$$

and the order of convergence is at least 1. Further, for the 'Riesz' data functions R_y, $y \in K$, we have

$$|R_y(x) - (\mathcal{I}_X R_y)(x)| \le k_G (h_{X,K})^2, \quad x, y \in K,$$

i.e. a 'doubled' estimate of at least 2 for the rate of convergence.

Since the weight function has property W2 for κ iff $\kappa + 1/2 < n$ we can choose $\kappa = 1$ iff $n \ge 2$. Then by Theorem 129, $k_G = (2\pi)^{-d/4} \sqrt{-\left(\nabla^2 G\right)(0)}$ and since

$$\begin{aligned}
\left(\nabla^2 G\right)(0) = \sum_{k=1}^{d} \left(D_k^2 G\right)(0) &= \sum_{k=1}^{d} D_k^2 \left(G_1(x_1) G_1(x_2) \ldots G_1(x_d)\right)(0) \\
&= \sum_{k=1}^{d} G_1(0)^{d-1} D^2 G_1(0) \\
&= G_1(0)^{d-1} D^2 G_1(0)\, d,
\end{aligned}$$

we get the formula

$$k_G = (2\pi)^{-d/4} \sqrt{-G_1(0)^{d-1} D^2 G_1(0)} \sqrt{d}. \tag{3.56}$$

3.6.2 Convergence estimates assuming unisolvent data sets when $\kappa \geq 1$

In fact comparison of Theorem 7, Chapter 1 for the extended B-splines with Theorem 98 of this document shows that for parameters n and l the smoothness parameters κ can take an identical sets of values when $\int_{\mathbb{R}^1} |\xi|^{2n-1} q(\xi)\, d\xi < \infty$, since $q \in L^1$.

We now require the following convergence estimate from Chapter 2:

Theorem 130 *(Copy of Theorem 74) Let w be a weight function with property W2 for integer parameter $\kappa \geq 1$ and let G be the corresponding basis function. Set $m = \lfloor \kappa \rfloor$.*

Suppose $\mathcal{I}_X f$ is the minimal norm interpolant of the data function $f \in X_w^0$ on the independent data set X contained in the data region Ω. We use the notation and assumptions of Lemma 72 of Chapter 2 which means assuming that X is m-unisolvent and Ω is a bounded, open, connected set whose boundary satisfies the cone condition.

Now set $k_G = \frac{d^{m/2}}{(2\pi)^{d/2} m!} (c_{\Omega,\kappa})^m K'_{\Omega,\kappa} \max\limits_{|\beta|=m} \left| D^{2\beta} G(0) \right|$. Then there exists $h_{\Omega,\kappa} > 0$ such that

$$|f(x) - (\mathcal{I}_X f)(x)| \leq k_G \|f\|_{w,0} (h_{X,\Omega})^m, \quad x \in \overline{\Omega}, \tag{3.57}$$

when $h_{X,\Omega} = \sup\limits_{\omega \in \Omega} \operatorname{dist}(\omega, X) < h_{\Omega,\kappa}$ i.e. the order of convergence is at least m.

The constants $c_{\Omega,m}$, $K'_{\Omega,m}$ and $h_{\Omega,m}$ only depend on Ω, m and d. In terms of the integrals which define weight property W2 we have

$$\max_{|\beta|=m} \left| D^{2\beta} G(0) \right| \leq (2\pi)^{-\frac{d}{2}} \int \frac{|\xi|^{2m} d\xi}{w(\xi)}.$$

We now show that the order of convergence of the minimal norm interpolant to an arbitrary data function is at least $n-1$:

Corollary 131 *Suppose the weight function is a central difference weight function with parameters $2 \leq n \leq l$. Then the order of convergence of the minimal norm interpolant to an arbitrary data function is at least $n-1$.*

Proof. Theorem 98 implies that $m = \lfloor \kappa \rfloor = n-1$ and Theorem 130 implies that the order of convergence is at least $n-1$. ∎

3.6.3 Summary table

Interpolant convergence to data function - central difference weight function.					
Estimate name	Unisolv. data?	Parameter constraints	Converg. order	$(2\pi)^{d/4} k_G$	h_G
Type 1	no	q b'nded if $n=1$ else $n \geq 2$	1/2	$\sqrt{2G_1(0)^{d-1} \|DG_1\|_\infty \sqrt[4]{d}}$	∞
Type 2	no	$n \geq 2$	1	$\sqrt{-G_1(0)^{d-1} D^2 G_1(0) \sqrt{d}}$	∞
Unisolvent	yes	$n \geq 2$	$n-1$	-	∞

TABLE 3.1.

4

The Exact smoother

4.1 Introduction

We call this well-known basis function smoother the *Exact smoother* because the smoother studied in the next chapter approximates it. We will assume the basis function is real-valued. The Exact smoother minimizes the functional

$$\rho \left\| f \right\|_{w,0}^2 + \frac{1}{N} \sum_{i=1}^{N} \left| f(x^{(i)}) - y_i \right|^2, \quad f \in X_w^0,$$

over the data (function) space X_w^0 where $\rho > 0$ is termed the smoothing coefficient. It is shown that this problem, like the interpolation problem, has a unique basis function solution in the space $W_{G,X}$. The finite dimensionality of the solution allows us to derive matrix equations for the coefficients α_i of the data-translated basis functions.

Three estimates are derived for the order of the pointwise error of the Exact smoother w.r.t. its data function: two estimates do not assume unisolvent X data and one does. When $\rho = 0$ these estimates correspond to interpolation results of Subsection 2.5.2 and the Exact smoother convergence orders and the constants are the same as those for the interpolation case which are given in the interpolation tables 1, 2, 3 and 4.

$\boxed{\text{Non-unisolvent data: Type 1 error estimates}}$ When the weight function has property W2 for $\kappa \geq 0$ it is assumed that the basis function satisfies an inequality of the form 1 and that the data region is a closed, bounded, infinite set K. In this case it is shown in Corollary 153 that the Exact smoother s_e of the data function f satisfies the error estimate

$$\left| f(x) - s_e(x) \right| \leq \left\| f \right\|_{w,0} \left(\sqrt{\rho N} + k_G \left(h_{X,K} \right)^s \right), \quad x \in K,$$

when $h_{X,K} = \max_{x \in K} \operatorname{dist}(x, X) \leq h_G$ and $k_G = (2\pi)^{-d/4} \sqrt{2C_G}$.

$\boxed{\text{Non-unisolvent data: Type 2 error estimates}}$ If it only assumed that $\kappa \geq 1$ then it is shown in Theorem 156 that

$$\left| f(x) - s_e(x) \right| \leq \left\| f \right\|_{w,0} \left(\sqrt{\rho N} + k_G h_{X,K} \right), \quad x \in K,$$

when $h_{X,K} < \infty$ and $k_G = (2\pi)^{-d/4} \sqrt{-\left(\nabla^2 G\right)(0)} \sqrt{d}$.

$\boxed{\text{Unisolvent data error estimates}}$ On the other hand, suppose the weight function has property W2 for some parameter $\kappa \geq 1$, the independent data X contains a unisolvent set of order $\lfloor \kappa \rfloor$ and X is

contained in an open, bounded data region Ω. Then using results from the Lagrange theory of polynomial interpolation we show in Theorem 160 that there exist constants $K'_{\Omega,m}, k_G > 0$ such that

$$\left| f(x) - s_e(x) \right| \leq \|f\|_{w,0} \left(K'_{\Omega,m} \sqrt{\rho N} + k_G \left(h_{X,\Omega} \right)^{\lfloor \kappa \rfloor} \right), \quad x \in \overline{\Omega},$$

when $h_{X,\Omega} = \sup\limits_{x \in \Omega} \operatorname{dist}(x, X) \leq h_G$.

These theoretical results will be illustrated using the weight function examples from the interpolation chapter, namely the radial *shifted thin-plate splines*, *Gaussian* and *Sobolev splines* and the tensor product *extended B-splines*. We will also use the *central difference* weight functions from Chapter 3.

Numerical results are only presented for the non-unisolvent data cases. Numeric experiments are carried out using the same 1-dimensional B-splines and data functions that were used for the interpolants. We restrict ourselves to one dimension so that the data density parameters $h_{X,\Omega}$ and $h_{X,K}$ can be easily calculated. Also note that in one dimension part 2 of Theorem 214 implies that any set of m distinct points is m-unisolvent.

No numerical experiments will be carried out for the unisolvent data case because I did not want to numerically estimate the parameters related to the Lagrange polynomial interpolation. However satisfactory easy estimates **may** be possible.

4.2 The Exact Smoothing problem

The Exact smoothing problem will be defined as a variational minimization problem within the space of continuous functions X_w^0. The functional is constructed using a real smoothing parameter $\rho > 0$ and a set of distinct scattered data points $\left\{ \left(x^{(k)}, y_k \right) \right\}_{k=1}^N$, $x^{(k)} \in \mathbb{R}^d$ and $y_k \in \mathbb{C}$. The functional to be minimized is

$$J_e[f] = \rho \|f\|_{w,0}^2 + \frac{1}{N} \sum_{k=1}^N \left| f(x^{(k)}) - y_k \right|^2, \quad f \in X_w^0. \tag{4.1}$$

The independent data will be specified by $X = \left(x^{(k)} \right)_{k=1}^N$ and the dependent data by $y = (y_k)_{k=1}^N$. The Exact smoothing problem is now stated as:

The Exact smoothing problem
Minimize the smoothing functional $J_e[f]$ for $f \in X_w^0$.

(4.2)

Here the first component of the functional $\rho \|f\|_{w,0}^2$ is the *global smoothing component* and the second component is the *localizing least squares component*. As $\rho \to 0^+$ it will be shown that the matrix equation representing this problem becomes that of the minimal norm interpolation problem 2.4. As the value of ρ increases numerical experiments show the degree of smoothing increases and the correlation between data and the smoothing function decreases.

Using Hilbert space techniques I will now do the following:

1. Show there exists a unique Exact smoother.
2. Show the smoother is a basis function smoother i.e. show that it lies in $W_{G,X}$.
3. Construct a matrix equation for the coefficients of the data-translated basis functions $G\left(\cdot - x^{(k)} \right)$.

The proofs will be carried out within a Hilbert space framework by formulating the smoothing functional J_e in terms of a norm. $\|\cdot\|_V$ on the Hilbert product space $V = X_w^0 \otimes \mathbb{C}^N$. To this end I first introduce the following definitions.

Definition 132 ***The space V and the operator $\mathcal{L}_X$***

1. *Let $V = X_w^0 \otimes \mathbb{C}^N$ be the Hilbert product space with norm $\|\cdot\|_V$ and inner product $(\cdot,\cdot)_V$ given by*

$$\left((u_1, u_2), (v_1, v_2) \right)_V = \rho \, (u_1, v_1)_{w,0} + \frac{1}{N} (u_2, v_2)_{\mathbb{C}^N}.$$

2. *Let the operator* $\mathcal{L}_X : \mathbb{C}^N \to V$ *be defined by* $\mathcal{L}_X f = \left(f, \widetilde{\mathcal{E}}_X f \right)$ *where* $\widetilde{\mathcal{E}}_X$ *is the vector-valued evaluation operator* $\widetilde{\mathcal{E}}_X f = \left(f\left(x^{(k)} \right) \right)$ *of Definition 51.*

Remark 133 *The definitions above were constructed so I have the following result: If* $y = (y_i)$ *is the (complex) dependent data given in the smoothing problem and if I let* $\varsigma = (0, y)$ *then for* $f \in X_w^0$ *the Exact smoothing functional can be rewritten in terms of the norm* $\|\cdot\|_V$ *as*

$$J_e[f] = \rho \|f\|_{w,0}^2 + \frac{1}{N} \left| \widetilde{\mathcal{E}}_X f - y \right|_{\mathbb{C}^N}^2 = \left\| \left(f, \widetilde{\mathcal{E}}_X f - y \right) \right\|_V^2 = \left\| \left(f, \widetilde{\mathcal{E}}_X f \right) - (0, y) \right\|_V^2 = \|\mathcal{L}_X f - \varsigma\|_V^2 .$$

The Exact smoother is now the (unique) orthogonal projection of the constant $(0, y)$ *onto the infinite dimensional subspace* $\mathcal{L}_X \left(X_w^0 \right)$.

I will need the following properties of the operators $\mathcal{L}_X$, $\mathcal{L}_X^*$ and $\mathcal{L}_X^* \mathcal{L}_X$.

Theorem 134 *The operator* $\mathcal{L}_X$ *has the following properties:*

1. $\|\mathcal{L}_X f\|_V$ *and* $\|f\|_{w,0}$ *are equivalent norms on* X_w^0.
2. $\mathcal{L}_X : X_w^0 \to V$ *is continuous, 1-1 and has closed range.*
3. $\mathcal{L}_X^* : V \to X_w^0$ *and if* $u = (u_1, u_2) \in V$ *then*

$$\mathcal{L}_X^* u = \rho u_1 + \frac{1}{N} \widetilde{\mathcal{E}}_X^* u_2.$$

4. $\mathcal{L}_X^* \mathcal{L}_X : X_w^0 \to X_w^0$ *and* $\mathcal{L}_X^* \mathcal{L}_X : W_{G,X} \to W_{G,X}$ *and*

$$\mathcal{L}_X^* \mathcal{L}_X = \rho I + \frac{1}{N} \widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X .$$

5. $\mathcal{L}_X^* \mathcal{L}_X$ *is continuous and 1-1 on* X_w^0.

Proof. Part 1. By part 3 of Theorem 52, $\left\| \widetilde{\mathcal{E}}_X \right\| = \|R_{X,X}\|$ so

$$\begin{aligned} \|\mathcal{L}_X f\|_V^2 = \left\| \left(f, \widetilde{\mathcal{E}}_X f \right) \right\|_V^2 = \rho \|f\|_{w,0}^2 + \frac{1}{N} \left| \widetilde{\mathcal{E}}_X f \right|_{\mathbb{C}^N}^2 &\le \rho \|f\|_{w,0}^2 + \frac{1}{N} \|R_{X,X}\|^2 \|f\|_{w,0}^2 \\ &= \left(\rho + \frac{1}{N} \|R_{X,X}\|^2 \right) \|f\|_{w,0}^2 . \end{aligned}$$

Further, since $\rho > 0$

$$\|f\|_{w,0}^2 \le \left(\min\left\{ \rho, \frac{1}{N} \right\} \right)^{-1} \left(\rho \|f\|_{w,0}^2 + \frac{1}{N} \left| \widetilde{\mathcal{E}}_X f \right|_{\mathbb{C}^N}^2 \right) = \left(\min\left\{ \rho, \frac{1}{N} \right\} \right)^{-1} \|\mathcal{L}_X f\|_V^2 .$$

Part 2. $\mathcal{L}_X : X_w^0 \to V$ is continuous since $\widetilde{\mathcal{E}}_X$ is continuous by part 1 of Theorem 52. From part 1 there exists a constant C_1 such that $\|f\|_{w,0} \le C_1 \|\mathcal{L}_X f\|_V$ and hence the range of $\mathcal{L}_X$ is closed and $\mathcal{L}_X$ is 1-1.

Part 3. Since $\mathcal{L}_X$ is a continuous operator the adjoint $\mathcal{L}_X^* : V \to X_w^0$ exists and is continuous. Further, if $f \in X_w^0$ and $u = (u_1, u_2) \in V$

$$\begin{aligned} (\mathcal{L}_X f, u)_V = \left(\left(f, \widetilde{\mathcal{E}}_X f \right), (u_1, u_2) \right)_V &= \rho\, (f, u_1)_{w,0} + \frac{1}{N} \left(\widetilde{\mathcal{E}}_X f, u_2 \right) \\ &= \rho\, (f, u_1)_{w,0} + \frac{1}{N} \left(f, \widetilde{\mathcal{E}}_X^* u_2 \right)_{w,0} \\ &= \left(f, \rho u_1 + \frac{1}{N} \widetilde{\mathcal{E}}_X^* u_2 \right)_{w,0} . \end{aligned}$$

Hence $\mathcal{L}_X^* u = \rho u_1 + \frac{1}{N}\widetilde{\mathcal{E}}_X^* u_2$, where $u = (u_1, u_2) \in V$.

Part 4. If $f \in X_w^0$ then

$$\mathcal{L}_X^* \mathcal{L}_X f = \rho \left(\mathcal{L}_X f\right)_1 + \frac{1}{N}\widetilde{\mathcal{E}}_X^* \left(\mathcal{L}_X f\right)_2 = \rho \left(f, \widetilde{\mathcal{E}}_X f\right)_1 + \frac{1}{N}\widetilde{\mathcal{E}}_X^* \left(f, \widetilde{\mathcal{E}}_X f\right)_2 = \rho f + \frac{1}{N}\widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X f.$$

By part 2 of Theorem 52 $\widetilde{\mathcal{E}}_X^* : \mathbb{C}^N \to W_{G,X}$ and it follows that $\mathcal{L}_X^* \mathcal{L}_X : X_w^0 \to X_w^0$ and $\mathcal{L}_X^* \mathcal{L}_X : W_{G,X} \to W_{G,X}$.

Part 5. Clearly $\mathcal{L}_X^* \mathcal{L}_X$ is continuous. Now suppose $\mathcal{L}_X^* \mathcal{L}_X f = 0$ for $f \in X_w^0$. Then $0 = \left(\mathcal{L}_X^* \mathcal{L}_X f, f\right)_{w,0} = \left(\mathcal{L}_X f, \mathcal{L}_X f\right)_{w,0} = \left\|\mathcal{L}_X f\right\|_V^2$ so that $\mathcal{L}_X f = 0$ and hence $f = 0$ since $\mathcal{L}_X$ is 1-1 by part 2. Therefore $\mathcal{L}_X^* \mathcal{L}_X$ is 1-1. ■

I next show that the Exact smoother, like the minimum norm interpolant of Chapter 2, has the nice property of being in the N-dimensional basis function space

$$W_{G,X} = \left\{ \sum_{i=1}^{N} \alpha_i G\left(\cdot - x^{(i)}\right) : \alpha_i \in \mathbb{C} \right\},$$

introduced in Definition 46. It was shown in Theorem 45 that the data-translated functions $G\left(\cdot - x^{(i)}\right)$ are linearly independent so $\dim W_{G,X} = N$.

Theorem 135 *Fix $y \in \mathbb{C}^N$ and let $\varsigma = (0, y) \in V$. Then there exists a unique solution to the Exact smoothing problem 4.2 in X_w^0. Denote this solution by s_e. This solution has the following properties:*

1. $\left\|\mathcal{L}_X s_e - \varsigma\right\|_V < \left\|\mathcal{L}_X f - \varsigma\right\|_V$, *for all $f \in X_w^0 - \{s_e\}$.*
2. $\left(\mathcal{L}_X s_e - \varsigma, \mathcal{L}_X \left(s_e - f\right)\right)_V = 0$, *for all $f \in X_w^0$.*
3. $\left\|\mathcal{L}_X s_e - \varsigma\right\|_V^2 + \left\|\mathcal{L}_X \left(s_e - f\right)\right\|_V^2 = \left\|\mathcal{L}_X f - \varsigma\right\|_V^2$, *for all $f \in X_w^0$.*

   The last equality is equivalent to the equality of part 2.
4. $s_e = \frac{1}{N}\left(\mathcal{L}_X^* \mathcal{L}_X\right)^{-1} \widetilde{\mathcal{E}}_X^* y$.
5. $s_e \in W_{G,X}$ *and*

$$s_e = \frac{1}{\rho N} \sum_{k=1}^{N} \left(y_k - s_e(x^{(k)})\right) R_{x^{(k)}}. \tag{4.3}$$

Proof. Parts 1, 2, 3. Because $\mathcal{L}_X(X_w^0)$ is closed it is well known there exists a unique element s_e of the hyperplane $\mathcal{L}_X(X_w^0)$ which minimizes the distance between the hyperplane $\mathcal{L}_X(X_w^0)$ and the point ς. From simple geometric considerations I follows that s_e satisfies conditions 1, 2 and 3 of this theorem.

Part 4. From part 2

$$0 = \left(\mathcal{L}_X s_e - \varsigma, \mathcal{L}_X s_e - \mathcal{L}_X f\right)_V = \left(\mathcal{L}_X s_e - \varsigma, \mathcal{L}_X \left(s_e - f\right)\right)_V = \left(\mathcal{L}_X^* \left(\mathcal{L}_X s_e - \varsigma\right), s_e - f\right)_{w,0},$$

for all $f \in X_w^0$. Thus

$$\mathcal{L}_X^* \left(\mathcal{L}_X s_e - \varsigma\right) = 0. \tag{4.4}$$

Hence

$$\mathcal{L}_X^* \mathcal{L}_X s_e = \mathcal{L}_X^* \varsigma = \mathcal{L}_X^* (0, y) = \frac{1}{N}\widetilde{\mathcal{E}}_X^* y,$$

and since by part 5 of Theorem 134 $\mathcal{L}_X^* \mathcal{L}_X$ is one-to-one, it follows that $\mathcal{L}_X^* \mathcal{L}_X$ has an inverse and so $s_e = \frac{1}{N}\left(\mathcal{L}_X^* \mathcal{L}_X\right)^{-1} \widetilde{\mathcal{E}}_X^* y$.

Part 5. Starting with equation 4.4

$$\begin{aligned} 0 = \mathcal{L}_X^* \left(\mathcal{L}_X s_e - \varsigma\right) = \mathcal{L}_X^* \mathcal{L}_X s_e - \frac{1}{N}\widetilde{\mathcal{E}}_X^* y &= \rho s_e + \frac{1}{N}\widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X s_e - \frac{1}{N}\widetilde{\mathcal{E}}_X^* y \\ &= \rho s_e + \frac{1}{N}\widetilde{\mathcal{E}}_X^* \left(\widetilde{\mathcal{E}}_X s_e - y\right) \\ &= \rho s_e - \frac{1}{N} \sum_{k=1}^{N} \left(s_e(x^{(k)}) - y_k\right) R_{x^{(k)}}, \end{aligned}$$

which proves 4.3. Finally, from the introduction to this paper $R_{x^{(k)}} = (2\pi)^{-d/2} G\left(\cdot - x^{(k)}\right)$ and so I have $s_e \in W_{G,X}$. ■

In a manner analogous to the case of the minimal norm interpolant, part 4 of the last theorem allows the definition of a mapping between a data function and it's corresponding Exact smoother.

Definition 136 ***Data functions and the Exact smoother operator*** $\mathcal{S}_X^e : X_w^0 \to W_{G,X}$.

Given an independent data set X, I shall assume that each member of X_w^0 can act as a legitimate data function g and generate the data vector $\widetilde{\mathcal{E}}_X g$ - see item 4 of Section 4.8.

The equation of part 4 of Theorem 135 enables us to define the continuous linear mapping $\mathcal{S}_X^e : X_w^0 \to W_{G,X}$ from the data functions to the corresponding Exact smoother given by

$$\mathcal{S}_X^e g = \frac{1}{N} \left(\mathcal{L}_X^* \mathcal{L}_X\right)^{-1} \widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X g, \quad g \in X_w^0. \tag{4.5}$$

It was shown in Theorem 134 that $\mathcal{L}_X^* \mathcal{L}_X : X_w^0 \to X_w^0$ is continuous and 1-1. I now prove $\mathcal{L}_X^* \mathcal{L}_X$ is onto and hence a homeomorphism. I also prove some important properties of the Exact smoother mapping $\mathcal{S}_X^e$.

Corollary 137 *$\mathcal{L}_X^* \mathcal{L}_X : X_w^0 \to X_w^0$ is a homeomorphism and $\mathcal{L}_X^* \mathcal{L}_X : W_{G,X} \to W_{G,X}$ is a homeomorphism.*

Regarding the properties of $\mathcal{S}_X^e$:

1. $\mathcal{S}_X^e g = g - \rho \left(\mathcal{L}_X^* \mathcal{L}_X\right)^{-1} g, \quad g \in X_w^0,$
2. *$\mathcal{S}_X^e$ is continuous but not a projection,*
3. *$\mathcal{S}_X^e : X_w^0 \to W_{G,X}$ is onto,*
4. $\operatorname{null} \mathcal{S}_X^e = W_{G,X}^{\perp}$,
5. *$\mathcal{S}_X^e$ is self-adjoint.*

Proof. Suppose $g \in X_w^0$ and let $s_e = \mathcal{S}_X^e g$. From part 5 of Theorem 134 I know that $\mathcal{L}_X^* \mathcal{L}_X$ is 1-1. From part 4 of Theorem 135 $\mathcal{L}_X^* \mathcal{L}_X s_e = \frac{1}{N} \widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X g$ and from part 4 of Theorem 134, $\mathcal{L}_X^* \mathcal{L}_X g = \rho g + \frac{1}{N} \widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X g$ so that $\mathcal{L}_X^* \mathcal{L}_X g = \rho g + \mathcal{L}_X^* \mathcal{L}_X s_e$. Clearly this equation implies part 1 and the equation

$$g = \mathcal{L}_X^* \mathcal{L}_X \left(\frac{g - s_e}{\rho}\right), \qquad g \in X_w^0. \tag{4.6}$$

This latter equation proves $g \in \operatorname{range} \mathcal{L}_X^* \mathcal{L}_X$. Thus $\mathcal{L}_X^* \mathcal{L}_X : X_w^0 \to X_w^0$ is onto and by the Open Mapping theorem $\mathcal{L}_X^* \mathcal{L}_X$ is a homeomorphism. If $g \in W_{G,X}$ then 4.6 implies $\mathcal{L}_X^* \mathcal{L}_X : W_{G,X} \to W_{G,X}$ is onto and so $\mathcal{L}_X^* \mathcal{L}_X : W_{G,X} \to W_{G,X}$ is also a homeomorphism. This fact and part 1 clearly imply that $\mathcal{S}_X^e$ is not a projection.

Further, in Theorem 52 it was shown that $\widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X : X_w^0 \to W_{G,X}$ is onto and $\operatorname{null} \widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X = W_{G,X}^{\perp}$. The stated properties of $\mathcal{S}_X^e$ now follow from 4.5 and the fact that $\widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X$ and $\mathcal{L}_X^* \mathcal{L}_X$ are self-adjoint. ■

4.3 Matrix equations for the Exact smoother

Part 5 of Theorem 135 means that the Exact smoother lies in the finite dimensional space $W_{G,X}$ and this enables matrix equations to be derived for the coefficients of the *data-translated* basis functions $G\left(x - x^{(k)}\right)$. These matrix equations can be written in terms of the *basis function matrix* $G_{X,X} = \left(G\left(x^{(i)} - x^{(j)}\right)\right)$ or the *reproducing kernel matrix* $R_{X,X} = \left(R_{x^{(j)}}\left(x^{(i)}\right)\right)$ where $R_{X,X} = (2\pi)^{-\frac{d}{2}} G_{X,X}$. The relevant properties of $G_{X,X}$ and $R_{X,X}$ were noted in the Introduction to this paper. $R_y(x) \neq G(x - y)$ because the Fourier transform I have selected is $\widehat{f}(\xi) = (2\pi)^{-d/2} \int e^{-ix\xi} f(x) dx$. For numeric algorithms the use of $G_{X,X}$ might be preferred because G is specified directly and for analysis the use of $R_{X,X}$ may be preferred because of the analytic properties of R_x.

Theorem 138 *Fix $y \in \mathbb{C}^N$ and let $\varsigma = (0, y) \in V$. Then by Theorem 135 there exists a unique solution to the Exact smoothing problem. I denote this solution by s and note that here, and in the sequel, I will sometimes use the more compact notation $s_X = \widetilde{\mathcal{E}}_X s$. The solution s has the following properties:*

1. *s_X satisfies the matrix equation*

$$\left(N\rho I + R_{X,X}\right) s_X = R_{X,X} y, \tag{4.7}$$

where $R_{X,X} = (2\pi)^{-\frac{d}{2}} G_{X,X}$.

2. *The matrix $N\rho I + R_{X,X}$ is real, symmetric, positive definite over $\mathbb{C}^N$ and regular.*

3. *$s(x) = \sum\limits_{k=1}^{N} \alpha_k R_{x^{(k)}}(x)$ where $\alpha = (\alpha_k)$ satisfies*

$$\left(N\rho I + R_{X,X}\right) \alpha = y, \tag{4.8}$$

and $\alpha = \frac{1}{N\rho}(y - s_X)$ when $\rho > 0$.

The next part is convenient for numerical purposes.

4. *$s(x) = \sum\limits_{k=1}^{N} \beta_k G\left(x - x^{(k)}\right)$ where $\beta = (\beta_k)$ satisfies*

$$\left((2\pi)^{\frac{d}{2}} N\rho I + G_{X,X}\right) \beta = y.$$

Proof. Part 1 Part 4 of Theorem 135 and part 4 of Theorem 134 imply

$$\mathcal{L}_X^* \mathcal{L}_X s = \rho I s + \frac{1}{N} \widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X s = \frac{1}{N} \widetilde{\mathcal{E}}_X^* y. \tag{4.9}$$

From Theorem 52 I have $\widetilde{\mathcal{E}}_X \widetilde{\mathcal{E}}_X^* = R_{X,X}$ so that applying the operator $(2\pi)^{d/2} N \widetilde{\mathcal{E}}_X$ to 4.9 gives 4.7.

Part 2 From the Introduction $R_{X,X} = (2\pi)^{-\frac{d}{2}} G_{X,X}$ is Hermitian and positive definite over $\mathbb{C}^N$. Since this paper assumes the basis function G is real valued, $R_{X,X}$ must be symmetric and thus $N\rho I + R_{X,X}$ is symmetric. Now suppose $\alpha \in \mathbb{C}^N$. Then

$$\alpha^T \left(N\rho I + R_{X,X}\right) \overline{\alpha} = N\rho \left|\alpha\right|^2 + \alpha^T R_{X,X} \overline{\alpha},$$

and since $R_{X,X}$ is positive definite over $\mathbb{C}^N$, $\alpha^T \left(N\rho I + R_{X,X}\right) \overline{\alpha} = 0$ implies $\alpha = 0$. Thus $N\rho I + R_{X,X}$ is positive definite over $\mathbb{C}^N$ and consequently must be regular.

Part 3 From part 5 of Theorem 135, $s(x) = \sum\limits_{k=1}^{N} \alpha_k R\left(x - x^{(k)}\right)$ where $\alpha_k = \frac{y_k - s\left(x^{(k)}\right)}{N\rho}$ i.e. $\alpha = (\alpha_k) = \frac{y - s_X}{N\rho}$. Thus $s_X = R_{X,X}\alpha$ and substituting for s_X in 4.7 we get
$\left(N\rho I + R_{X,X}\right) R_{X,X} \alpha = R_{X,X} y$ i.e. matrix equation 4.8 since $R_{X,X}$ is regular.

Part 4 A straightforward consequence of part 3. ■

The next corollary will prove very useful.

Corollary 139 *If $g \in X_w^0$ is a (complex) data function then the Exact smoother operator $\mathcal{S}_X^e$ satisfies*

$$\mathcal{S}_X^e g = \widetilde{\mathcal{E}}_X^* \left(N\rho I + R_{X,X}\right)^{-1} \widetilde{\mathcal{E}}_X g, \tag{4.10}$$

and

$$\left(\mathcal{S}_X^e g\right)(x) = \left(\widetilde{\mathcal{E}}_X R_x\right)^T \left(N\rho I + R_{X,X}\right)^{-1} \widetilde{\mathcal{E}}_X g. \tag{4.11}$$

Proof. The data vector is $\widetilde{\mathcal{E}}_X g$ and from part 3 of the last theorem $\left(\mathcal{S}_X^e g\right)(x) = \sum\limits_{k=1}^{N} \alpha_k R_{x^{(k)}}(x)$ where $\alpha = (\alpha_k)$ is given by $\alpha = \left(N\rho I + R_{X,X}\right)^{-1} \widetilde{\mathcal{E}}_X g$. Now making use of the results concerning $\widetilde{\mathcal{E}}_X$ and $\widetilde{\mathcal{E}}_X^*$ in parts 1 and 2 of Theorem 52 we have

$$s = \mathcal{S}_X^e g = \sum_{k=1}^{N} \alpha_k R_{x^{(k)}} = \widetilde{\mathcal{E}}_X^* \alpha = \widetilde{\mathcal{E}}_X^* \left(N\rho I + R_{X,X}\right)^{-1} \widetilde{\mathcal{E}}_X g,$$

and so, since the basis function is assumed to be real valued,

$$\begin{aligned} \mathcal{S}_X^e g(x) = \left(\widetilde{\mathcal{E}}_X^* \alpha, R_x\right)_{w,0} = \left(\alpha, \widetilde{\mathcal{E}}_X R_x\right)_{\mathbb{C}^N} &= \left(\left(N\rho I + R_{X,X}\right)^{-1} \widetilde{\mathcal{E}}_X g, \widetilde{\mathcal{E}}_X R_x\right)_{\mathbb{C}^N} \\ &= \left(\widetilde{\mathcal{E}}_X g\right)^T \left(N\rho I + R_{X,X}\right)^{-1} \widetilde{\mathcal{E}}_X R_x \\ &= \left(\widetilde{\mathcal{E}}_X R_x\right)^T \left(N\rho I + R_{X,X}\right)^{-1} \widetilde{\mathcal{E}}_X g. \end{aligned}$$

■

4.4 More properties of the Exact smoother

Theorem 140 *By Theorem 135 there exists a unique solution s to the Exact smoothing problem. This solution has the following properties:*

1. *For all $f \in X_w^0$*

$$\begin{aligned} \rho \|s\|_{w,0}^2 + \frac{1}{N}\sum_{k=1}^{N} \left|s\left(x^{(k)}\right) - y_k\right|^2 + \rho \|s - f\|_{w,0}^2 + \frac{1}{N}\sum_{k=1}^{N} \left|s\left(x^{(k)}\right) - f\left(x^{(k)}\right)\right|^2 & \\ = \rho \|f\|_{w,0}^2 + \frac{1}{N}\sum_{k=1}^{N} \left|f\left(x^{(k)}\right) - y_k\right|^2 . & \quad (4.12) \end{aligned}$$

2. *If f is the data function then $\|s\|_{w,0} \le \|f\|_{w,0}$ and $\|s - f\|_{w,0} \le \|f\|_{w,0}$.*
3. $\|s\|_{w,0}^2 = \frac{1}{\rho N} \operatorname{Re} \sum\limits_{k=1}^{N} \overline{s\left(x^{(k)}\right)} \left(y_k - s\left(x^{(k)}\right)\right)$.
4. $J_e[s] = \frac{1}{N} \operatorname{Re} \sum\limits_{k=1}^{N} \left(y_k - s\left(x^{(k)}\right)\right) \overline{y_k}$.

Proof. Part 1. From part 3 of Theorem 135

$$\|\mathcal{L}_X s - \varsigma\|_V^2 + \|\mathcal{L}_X (s - f)\|_V^2 = \|\mathcal{L}_X f - \varsigma\|_V^2 ,$$

where $\varsigma = (0, y)$. Part 1 follows from Remark 133 and the definition of the norm on the space V.

Part 2. If f is the data function then $f\left(x^{(k)}\right) = y_k$ and part 1 implies part 2.

Part 3. Substituting $f = 0$ in equation 4.12 of part 1 yields

$$\begin{aligned} 2\rho \|s\|_{w,0}^2 &= \frac{1}{N}\sum_{k=1}^{N} |y_k|^2 - \frac{1}{N}\sum_{k=1}^{N} \left|s\left(x^{(k)}\right) - y_k\right|^2 - \frac{1}{N}\sum_{k=1}^{N} \left|s\left(x^{(k)}\right)\right|^2 \\ &= \frac{1}{N}\sum_{k=1}^{N} \left(|y_k|^2 - \left|s\left(x^{(k)}\right) - y_k\right|^2 - \left|s\left(x^{(k)}\right)\right|^2\right) \\ &= \frac{1}{N}\sum_{k=1}^{N} \left(-2\overline{s\left(x^{(k)}\right)} s\left(x^{(k)}\right) + 2\operatorname{Re}\left(\overline{s\left(x^{(k)}\right)} y_k\right)\right) \\ &= \frac{2}{N} \operatorname{Re} \sum_{k=1}^{N} \overline{s\left(x^{(k)}\right)} \left(y_k - s\left(x^{(k)}\right)\right) . \end{aligned}$$

Part 4. By definition, $J_e[s] = \rho \|s\|_{w,0}^2 + \frac{1}{N}\sum\limits_{k=1}^{N} \left|s\left(x^{(k)}\right) - y_k\right|^2$.

Substituting for $\|s\|_{w,0}^2$ using the result of part 2 I get

$$\begin{aligned}
J_e[s] &= \rho \|s\|_{w,0}^2 + \frac{1}{N}\sum_{k=1}^{N} \left|s\left(x^{(k)}\right) - y_k\right|^2 \\
&= \frac{1}{N} \operatorname{Re} \sum_{k=1}^{N} \overline{s\left(x^{(k)}\right)} \left(y_k - s\left(x^{(k)}\right)\right) + \frac{1}{N}\sum_{k=1}^{N} \left|s\left(x^{(k)}\right) - y_k\right|^2 \\
&= \frac{1}{N}\sum_{k=1}^{N} \left(\begin{array}{c} \operatorname{Re}\left(\overline{s\left(x^{(k)}\right)} y_k - \overline{s\left(x^{(k)}\right)} s\left(x^{(k)}\right)\right) + \\ + \overline{s\left(x^{(k)}\right)} s\left(x^{(k)}\right) - 2\operatorname{Re} \overline{s\left(x^{(k)}\right)} y_k + \overline{y_k} y_k \end{array} \right) \\
&= \frac{1}{N} \operatorname{Re} \sum_{k=1}^{N} \left(-\overline{s\left(x^{(k)}\right)} y_k + \overline{y_k} y_k\right) \\
&= \frac{1}{N} \operatorname{Re} \sum_{k=1}^{N} \left(\overline{y_k} - \overline{s\left(x^{(k)}\right)}\right) y_k \\
&= \frac{1}{N} \operatorname{Re} \sum_{k=1}^{N} \left(y_k - s\left(x^{(k)}\right)\right) \overline{y_k}.
\end{aligned}$$

■

Remark 141 *Part 1 of the last theorem confirms that if s is the Exact smoother then $J_e[s] \le J_e[f]$ for all $f \in X_w^0$ i.e. that s minimizes the functional J_e.*

4.5 Semi-inner products and pointwise error estimates

Following the example of the interpolation seminorm $(f - \mathcal{I}_X f, g)_{w,0}$ of Definition 62 of 2 and it's sequel I will define the two semi-inner products $(\mathcal{S}_X^e f, g)_{w,0}$ and $(f - \mathcal{S}_X^e f, g)_{w,0}$ and use these to study the value and the error of the Exact smoother respectively. Of course these semi-inner products are useful because $(\mathcal{S}_X^e f, R_x)_{w,0} = (\mathcal{S}_X^e f)(x)$ and $(f - \mathcal{S}_X^e f, R_x)_{w,0} = (f - \mathcal{S}_X^e f)(x)$ and because of the power of Hilbert space theory. An outcome of this approach is the Exact smoother bound 4.15.

To define these smoother seminorms I will need the lemma:

Lemma 142 *Suppose $f \in X_w^0$, $\mathcal{S}_X^e$ is the Exact smoother mapping and $X = \left\{x^{(k)}\right\}_{k=1}^N$. Then:*

1. $\|\mathcal{S}_X^e f\|_{w,0}^2 = (\mathcal{S}_X^e f, f)_{w,0} - \rho N \left|(\rho N I + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X f\right|^2$.
2. $\|f - \mathcal{S}_X^e f\|_{w,0}^2 = (f - \mathcal{S}_X^e f, f)_{w,0} - \rho N \left|(\rho N I + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X f\right|^2$.

Proof. Part 1. I start with formula 4.10 for $\mathcal{S}_X^e f$.

$$\begin{aligned}
\|\mathcal{S}_X^e f\|_{w,0}^2 &= \left(\widetilde{\mathcal{E}}_X^* (\rho N I + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X f, \widetilde{\mathcal{E}}_X^* (\rho N I + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X f\right)_{w,0} \\
&= \left(\widetilde{\mathcal{E}}_X \widetilde{\mathcal{E}}_X^* (\rho N I + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X f, (\rho N I + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X f\right)_{\mathbb{C}^N} \\
&= \left(R_{X,X} (\rho N I + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X f, (\rho N I + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X f\right)_{\mathbb{C}^N},
\end{aligned}$$

since $\widetilde{\mathcal{E}}_X \widetilde{\mathcal{E}}_X^* = R_{X,X}$ but

$$\begin{aligned}
R_{X,X} (\rho N I + R_{X,X})^{-1} &= (\rho N I + R_{X,X}) (\rho N I + R_{X,X})^{-1} - \rho N (\rho N I + R_{X,X})^{-1} \\
&= I - \rho N (\rho N I + R_{X,X})^{-1},
\end{aligned}$$

so that

$$\begin{aligned}\|\mathcal{S}_X^e f\|_{w,0}^2 &= \left(\widetilde{\mathcal{E}}_X f, (\rho NI + R_{X,X})^{-1}\widetilde{\mathcal{E}}_X f\right) - \rho N\left((\rho NI + R_{X,X})^{-1}\widetilde{\mathcal{E}}_X f, (\rho NI + R_{X,X})^{-1}\widetilde{\mathcal{E}}_X f\right)\\ &= \left(\widetilde{\mathcal{E}}_X^* (\rho NI + R_{X,X})^{-1}\widetilde{\mathcal{E}}_X f, f\right)_{w,0} - \rho N\left|(\rho NI + R_{X,X})^{-1}\widetilde{\mathcal{E}}_X f\right|^2\\ &= (\mathcal{S}_X^e f, f)_{w,0} - \rho N\left|(\rho NI + R_{X,X})^{-1}\widetilde{\mathcal{E}}_X f\right|^2.\end{aligned}$$

Part 2. Substituting the equation for $\|\mathcal{S}_X^e f\|_{w,0}^2$ proved in part 1 yields

$$\begin{aligned}\|f - \mathcal{S}_X^e f\|_{w,0}^2 &= \|f\|_{w,0}^2 - 2(\mathcal{S}_X^e f, f)_{w,0} + \|\mathcal{S}_X^e f\|_{w,0}^2\\ &= \|f\|_{w,0}^2 - 2(\mathcal{S}_X^e f, f)_{w,0} + (\mathcal{S}_X^e f, f)_{w,0} - \rho N\left|(\rho NI + R_{X,X})^{-1}\widetilde{\mathcal{E}}_X f\right|^2\\ &= \|f\|_{w,0}^2 - (\mathcal{S}_X^e f, f)_{w,0} - \rho N\left|(\rho NI + R_{X,X})^{-1}\widetilde{\mathcal{E}}_X f\right|^2\\ &= (f - \mathcal{S}_X^e f, f)_{w,0} - \rho N\left|(\rho NI + R_{X,X})^{-1}\widetilde{\mathcal{E}}_X f\right|^2.\end{aligned}$$

■

Theorem 143 *The expressions $(\mathcal{S}_X^e f, g)_{w,0}$ and $(f - \mathcal{S}_X^e f, g)_{w,0}$ are both semi-inner products on X_w^0 with null spaces $W_{G,X}^{\perp}$.*

Proof. From part 1 of Lemma 142, $(\mathcal{S}_X^e f, f)_{w,0} \geq 0$ and $(\mathcal{S}_X^e f, f)_{w,0} = 0$ iff $(\rho NI + R_{X,X})^{-1}\widetilde{\mathcal{E}}_X f = 0$ iff $\widetilde{\mathcal{E}}_X f$ iff $f \in W_{G,X}^{\perp}$. Part 2 of Lemma 142 can similarly be applied to $(f - \mathcal{S}_X^e f, g)_{w,0}$. ■

I can now define our Approximate smoother seminorms on X_w^0.

Definition 144 ***Approximate smoother value seminorm*** *and* ***Approximate smoother error seminorm*** *on X_w^0.*

Define the Approximate smoother ***value*** *seminorm and semi-inner product by*

$$|f|_{sv}^2 = (\mathcal{S}_X^e f, f)_{w,0}, \qquad \langle f, g\rangle_{sv} = (\mathcal{S}_X^e f, g)_{w,0}.$$

The qualifier 'value' is used because $\langle f, R_x\rangle_{sv} = (\mathcal{S}_X^e f)(x)$.

Denote the Approximate smoother ***error*** *seminorm and semi-inner product by*

$$|f|_{se}^2 = (f - \mathcal{S}_X^e f, f)_{w,0}, \qquad \langle f, g\rangle_{se} = (f - \mathcal{S}_X^e f, g)_{w,0}.$$

The qualifier 'error' is used because $\langle f, R_x\rangle_{se} = (f - \mathcal{S}_X^e f)(x)$.

In the next theorem the pointwise error problem for an arbitrary function is reduced to that of considering the error of the smoother of R_x at x, where R_x is the Riesz representer of the evaluation functional $f \to f(x)$. This error will latter be estimated using the values of the basis function near the origin. This theorem also gives an error bound that is uniform on $\mathbb{R}^d$.

Theorem 145 *Since $\mathcal{S}_X^e f_d$ is the Exact smoother of the data function $f_d \in X_w^0$ and $\mathcal{S}_X^e R_x$ is the Exact smoother of the Riesz representer data function R_x, we have the bound*

$$|f_d(x) - \mathcal{S}_X^e f_d(x)| \leq \sqrt{(f_d - \mathcal{S}_X^e f_d, f_d)_{w,0}}\sqrt{R_x(x) - (\mathcal{S}_X^e R_x)(x)}, \quad x \in \mathbb{R}^d, \tag{4.13}$$

where $\sqrt{(f_d - \mathcal{S}_X^e f_d, f_d)_{w,0}} \leq \|f_d\|_{w,0}$, as well as the bound

$$|f_d(x) - \mathcal{S}_X^e f_d(x)| \leq \sqrt{(f_d - \mathcal{S}_X^e f_d, f_d)_{w,0}}\sqrt{R_0(0)}, \quad x \in \mathbb{R}^d. \tag{4.14}$$

Proof. Estimate 4.13 is just the Cauchy-Schwartz inequality $|\langle f_d, R_x\rangle_{se}| \leq |f_d|_{se}\, |R_x|_{se}$.

The second inequality of this theorem follows directly from the inequality $\|f_d - \mathcal{S}_X^e f_d\|_{w,0} \leq \|f_d\|_{w,0}$, proven in part 2 of Theorem 140. Thus

$$\begin{aligned} |f_d(x) - \mathcal{S}_X^e f_d(x)| &\leq \sqrt{(f_d - \mathcal{S}_X^e f_d, f_d)_{w,0}} \sqrt{R_x(x) - (\mathcal{S}_X^e R_x)(x)} \\ &\leq \sqrt{(f_d - \mathcal{S}_X^e f_d, f_d)_{w,0}} \sqrt{R_x(x)} \\ &= \sqrt{(f_d - \mathcal{S}_X^e f_d, f_d)_{w,0}} \sqrt{R_0(0)}. \end{aligned}$$

■

The last result means that convergence results about arbitrary data functions in X_w^0 can be proved by deriving error results for the Riesz representer function R_x at x.

I will now use the smoothing value seminorm to put bounds on the smoother that show how the smoother behaves for large values of the smoothing parameter ρ: it decreases to zero at at least the rate $1/\rho$.

Theorem 146 *If $f \in X_w^0$ and s_e is the Exact smoother of f with smoothing coefficient ρ then*

$$\begin{aligned} |s_e(x)| &\leq \begin{cases} \|f\|_{w,0}\sqrt{R_0(0)}, & \rho \leq R_0(0), \\ \|f\|_{w,0}\left(\sqrt{R_0(0)}\right)^3 \rho^{-1}, & \rho \geq R_0(0), \end{cases} \\ &= \|f\|_{w,0}\sqrt{R_0(0)} \min\left\{1, \frac{R_0(0)}{\rho}\right\}, \end{aligned} \tag{4.15}$$

where $R_0(0) = (2\pi)^{-\frac{d}{2}} G(0) = \|R_0\|_{w,0}^2$.

Proof. If $\mathcal{S}_X^e$ is the Exact smoother mapping then

$$|\mathcal{S}_X^e f(x)| = |(\mathcal{S}_X^e f, R_x)| \leq \|\mathcal{S}_X^e f\|_{w,0} \|R_x\|_{w,0} \leq \|f\|_{w,0} \|R_x\|_{w,0} = \|f\|_{w,0}\sqrt{R_0(0)}. \tag{4.16}$$

From part 1 of Lemma 142

$$\left|(\rho NI + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X f\right| < \frac{1}{\sqrt{\rho N}} \sqrt{(\mathcal{S}_X^e f, f)_{w,0}}, \quad f \in X_w^0,$$

and from part 3 of Theorem 52, $\left|\widetilde{\mathcal{E}}_X f\right| \leq \|f\|_{w,0}\sqrt{N}\sqrt{R_0(0)}$. Hence

$$\begin{aligned} (\mathcal{S}_X^e f, f)_{w,0} = \left(\widetilde{\mathcal{E}}_X f\right)^T (\rho NI + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X f &\leq \left|\widetilde{\mathcal{E}}_X f\right| \left|(\rho NI + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X f\right| \\ &\leq \|f\|_{w,0}\left(\sqrt{N}\sqrt{R_0(0)}\right) \frac{1}{\sqrt{\rho N}} \sqrt{(\mathcal{S}_X^e f, f)_{w,0}} \\ &= \|f\|_{w,0}\sqrt{R_0(0)} \frac{1}{\sqrt{\rho}} \sqrt{(\mathcal{S}_X^e f, f)_{w,0}}, \end{aligned}$$

so that $(\mathcal{S}_X^e f, f)_{w,0} \leq \|f\|_{w,0}^2 R_0(0)/\rho$. But by Theorem 145 $(\mathcal{S}_X^e f, g)_{w,0}$ is a semi-inner product and hence by the Cauchy-Schwartz theorem

$$\left|(\mathcal{S}_X^e f, g)_{w,0}\right| \leq \sqrt{(\mathcal{S}_X^e f, f)_{w,0}} \sqrt{(\mathcal{S}_X^e g, g)_{w,0}}, \quad f, g \in X_w^0,$$

which allows us to conclude that

$$|\mathcal{S}_X^e f(x)| = \left|(\mathcal{S}_X^e f, R_x)_{w,0}\right| \leq \sqrt{(\mathcal{S}_X^e f, f)_{w,0}} \sqrt{(\mathcal{S}_X^e R_x, R_x)_{w,0}} \leq \|f\|_{w,0}\left(\sqrt{R_0(0)}\right)^3 \frac{1}{\rho}.$$

which when combined with inequality 4.16 proves the first estimates of this theorem. ■

Remark 147 *These bounds on the smoother make more sense if we note that $|f(x)| \leq \|f\|_{w,0}\sqrt{R_0(0)}$.*

The goal of the next theorem is to prove the inequality of part 3 of this theorem. This inequality is proved using the identity of part 2 which is in turn derived by expanding the two terms $\|R_x - R_{x^{(k)}}\|_{w,0}^2 - \|\mathcal{S}_X^e (R_x - R_{x^{(k)}})\|_{w,0}^2$. This approach is motivated by the identity of part 2 of Theorem 64, namely, $R_x(x) - (\mathcal{I}_X R_x)(x) = \|R_x - R_{x^{(k)}}\|_{w,0}^2 - \|\mathcal{I}_X (R_x - R_{x^{(k)}})\|_{w,0}^2$, which was proved by expanding its right side. Here $\mathcal{I}_X$ is the interpolant mapping.

Theorem 148 *Recall that $|f|_{se}^2 = (\mathcal{S}_X^e f - f)(x)$ is the Exact smoother error seminorm introduced in Definition 144. Then for $x, y \in \mathbb{R}^d$:*

1.

$$\begin{aligned}\|\mathcal{S}_X^e (R_x - R_y)\|_{w,0}^2 = (\mathcal{S}_X^e R_x)(x) - 2(\mathcal{S}_X^e R_x)(y) + (\mathcal{S}_X^e R_y)(y) - \\ - \rho N \left|(\rho NI + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X (R_x - R_y)\right|^2.\end{aligned}$$

2.

$$|R_x - R_y|_{se}^2 = \|R_x - R_y\|_{w,0}^2 - \|\mathcal{S}_X^e (R_x - R_y)\|_{w,0}^2 - \rho N \left|(\rho NI + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X (R_x - R_y)\right|^2.$$

3. $|R_x|_{se} \le |R_y|_{se} + \|R_x - R_y\|_{w,0}$.

Proof. Part 1. Using the equation of part 1 of Lemma 142 with $f = R_x - R_y$

$$\begin{aligned}\|\mathcal{S}_X (R_x - R_y)\|_{w,0}^2 &= (\mathcal{S}_X^e (R_x - R_y), R_x - R_y)_{w,0} - \rho N \left|(\rho NI + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X (R_x - R_y)\right|^2 \\ &= (\mathcal{S}_X^e R_x, R_x)_{w,0} - 2(\mathcal{S}_X^e R_x, R_y)_{w,0} + (\mathcal{S}_X^e R_y, R_y)_{w,0} - \\ &\quad - \rho N \left|(\rho NI + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X (R_x - R_y)\right|^2 \\ &= (\mathcal{S}_X^e R_x)(x) - 2(\mathcal{S}_X^e R_x)(y) + (\mathcal{S}_X^e R_y)(y) - \\ &\quad - \rho N \left|(\rho NI + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X (R_x - R_y)\right|^2.\end{aligned}$$

Part 2. Since $\|R_x - R_y\|_{w,0}^2 = R_x(x) - 2R_x(y) + R_y(y)$, by part 1

$$\begin{aligned}\|R_x - R_y\|_{w,0}^2 - \|\mathcal{S}_X (R_x - R_y)\|_{w,0}^2 &= (R_x(x) - (\mathcal{S}_X^e R_x)(x)) - 2(R_x(y) - (\mathcal{S}_X^e R_x)(y)) + \\ &\quad + (R_y(y) - (\mathcal{S}_X^e R_y)(y)) + \\ &\quad + \rho N \left|(\rho NI + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X (R_x - R_y)\right|^2 \\ &= |R_x|_{se}^2 + |R_y|_{se}^2 - 2\langle R_x, R_y\rangle_{se} + \\ &\quad + \rho N \left|(\rho NI + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X (R_x - R_y)\right|^2 \\ &= |R_x - R_y|_{se}^2 + \rho N \left|(\rho NI + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X (R_x - R_y)\right|^2,\end{aligned}$$

and rearranging gives the result.

Part 3. From part 2, $|R_x - R_y|_{se} \le \|R_x - R_y\|_{w,0}$ and since $|R_x - R_y|_{se} \ge |R_x|_{se} - |R_y|_{se}$ we have part 3. ■

I now need the following lemma:

Lemma 149 *Suppose the weight function has property W2 for some $\kappa \ge 0$. Suppose $X = \{x^{(k)}\}_{k=1}^N$ is the independent data for the Exact smoothing operator $\mathcal{S}_X^e$. Then*

$$R_{x^{(j)}}\left(x^{(k)}\right) - (\mathcal{S}_X^e R_{x^{(j)}})\left(x^{(k)}\right) = \rho N \delta_{j,k} - (\rho N)^2 \left((\rho NI + R_{X,X})^{-1}\right)_{j,k}, \tag{4.17}$$

where $\left((\rho NI + R_{X,X})^{-1}\right)_{j,k}$ *is the* j,k *th element of* $(\rho NI + R_{X,X})^{-1}$. *Also*

$$0 < \left((\rho NI + R_{X,X})^{-1}\right)_{k,k} \leq \frac{1}{\rho N}.$$

Proof. Using equation 4.11 for $\left(\mathcal{S}_X^e R_{x^{(j)}}\right)\left(x^{(k)}\right)$ implies

$$\begin{aligned}
R_{x^{(j)}}\left(x^{(k)}\right) - \left(\mathcal{S}_X^e R_{x^{(j)}}\right)\left(x^{(k)}\right) &= R_{x^{(j)}}\left(x^{(k)}\right) - \left(\widetilde{\mathcal{E}}_X R_{x^{(k)}}\right)^T (\rho NI + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X R_{x^{(j)}} \\
&= R_{x^{(j)}}\left(x^{(k)}\right) - \left(\widetilde{\mathcal{E}}_X R_{x^{(j)}}\right)^T (\rho NI + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X R_{x^{(k)}} \\
&= R_{x^{(j)}}\left(x^{(k)}\right) - \left(\widetilde{\mathcal{E}}_X R_{x^{(j)}}\right)^T (\rho NI + R_{X,X})^{-1} \left(\rho N \mathbf{i}_k + \widetilde{\mathcal{E}}_X R_{x^{(k)}}\right) + \\
&\quad + \left(\widetilde{\mathcal{E}}_X R_{x^{(j)}}\right)^T (\rho NI + R_{X,X})^{-1} (\rho N \mathbf{i}_k) \\
&= R_{x^{(j)}}\left(x^{(k)}\right) - \left(\widetilde{\mathcal{E}}_X R_{x^{(j)}}\right)^T \mathbf{i}_k + \\
&\quad + \left(\widetilde{\mathcal{E}}_X R_{x^{(j)}}\right)^T (\rho NI + R_{X,X})^{-1} (\rho N \mathbf{i}_k) \\
&= \left(\widetilde{\mathcal{E}}_X R_{x^{(j)}}\right)^T (\rho NI + R_{X,X})^{-1} (\rho N \mathbf{i}_k) \\
&= \rho N \left(\widetilde{\mathcal{E}}_X R_{x^{(j)}}\right)^T (\rho NI + R_{X,X})^{-1} \mathbf{i}_k \\
&= (\rho N)\, \mathbf{i}_k^T (\rho NI + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X R_{x^{(j)}} \\
&= (\rho N)\, \mathbf{i}_k^T (\rho NI + R_{X,X})^{-1} \left(\rho N \mathbf{i}_j + \widetilde{\mathcal{E}}_X R_{x^{(j)}}\right) - \\
&\quad - (\rho N)\, \mathbf{i}_k^T (\rho NI + R_{X,X})^{-1} (\rho N \mathbf{i}_j) \\
&= \rho N \mathbf{i}_k^T \mathbf{i}_j - (\rho N)^2\, \mathbf{i}_k^T (\rho NI + R_{X,X})^{-1} \mathbf{i}_j \\
&= \rho N \delta_{j,k} - (\rho N)^2 \left((\rho NI + R_{X,X})^{-1}\right)_{j,k},
\end{aligned}$$

where $\left((\rho NI + R_{X,X})^{-1}\right)_{j,k}$ is element j,k of $(\rho NI + R_{X,X})^{-1}$.
Since $\rho NI + R_{X,X}$ is positive definite, $(\rho NI + R_{X,X})^{-1}$ is positive definite and thus $\mathbf{i}_k^T (\rho NI + R_{X,X})^{-1} \mathbf{i}_k > 0$.

Remark 150 *Equations 4.17 are equivalent to the matrix equation*

$$R_{X,X} - \left(\left(\mathcal{S}_X^e R_{x^{(j)}}\right)\left(x^{(i)}\right)\right) = \rho NI - (\rho N)^2 (\rho NI + R_{X,X})^{-1},$$

i.e.

$$\left(\langle R_{x^{(i)}}, R_{x^{(j)}} \rangle_{se}\right) = \rho NI - (\rho N)^2 (\rho NI + R_{X,X})^{-1}.$$

■

The next theorem shows how the error of the smoother of a function R_y is related to the behavior of the basis function near the origin, as well as to the number of data points and the smoothing coefficient.

Theorem 151 *Suppose* $\mathcal{S}_X^e$ *is the Exact smoother mapping with smoothing coefficient* ρ *and* $X = \left\{x^{(k)}\right\}_{k=1}^N$. *Then for each* $x^{(k)} \in X$

$$\sqrt{R_x(x) - \left(\mathcal{S}_X^e R_x\right)(x)} \leq \sqrt{\rho N - (\rho N)^2 \left((\rho NI + R_{X,X})^{-1}\right)_{k,k}} + \frac{\sqrt{2}}{(2\pi)^{\frac{d}{4}}} \sqrt{G(0) - G\left(x - x^{(k)}\right)},$$

where $\left((\rho NI + R_{X,X})^{-1}\right)_{k,k}$ *is the* k*th (positive) main diagonal element of the matrix* $(\rho NI + R_{X,X})^{-1}$.

Proof. In part 3 of Theorem 148 it was shown that $|R_x|_{se} \leq |R_y|_{se} + \|R_x - R_y\|_{w,0}$ where $|\cdot|_{se}$ is the smoother error seminorm. Setting $y = x^{(k)} \in X$ and using equation 4.17 with $j = k$ this inequality becomes

$$\begin{aligned} |R_x|_{se} = \sqrt{R_x(x) - (\mathcal{S}_X^e R_x)(x)} &\leq \sqrt{R_{x^{(k)}}\left(x^{(k)}\right) - \left(\mathcal{S}_X^e R_{x^{(k)}}\right)\left(x^{(k)}\right)} + \|R_x - R_{x^{(k)}}\|_{w,0} \\ &= \sqrt{\rho N - (\rho N)^2 \left(\left(\rho N I + R_{X,X}\right)^{-1}\right)_{k,k}} + \|R_x - R_{x^{(k)}}\|_{w,0}, \end{aligned}$$

and since $R_x(y) = (2\pi)^{-d/2} G(y - x)$

$$\begin{aligned} \|R_x - R_{x^{(k)}}\|_{w,0}^2 = \left(R_x - R_{x^{(k)}}, R_x - R_{x^{(k)}}\right)_{w,0} &= R_x(x) - 2R_x\left(x^{(k)}\right) + R_{x^{(k)}}\left(x^{(k)}\right) \\ &= \frac{2}{(2\pi)^{d/2}}\left(G(0) - G\left(x - x^{(k)}\right)\right), \end{aligned} \tag{4.18}$$

and our result follows. ■

4.6 Pointwise error estimates without assuming unisolvent data sets

In this section we will prove the Exact smoother analogues of the Type 1 and 2 interpolant error estimates of Subsection 2.5.2. Recall that for Type 1 estimates no restriction is placed on the weight function parameter κ but a local smoothness condition is imposed on the basis function near the origin. On the other hand Type 2 conditions only assume $\kappa \geq 1$ and we use the resultant basis function smoothness properties. The interpolation weight function examples will be used, augmented by the central difference weight functions.

4.6.1 Type 1 error estimates ($\kappa \geq 0$)

In the next two corollaries a smoothness condition is applied to the basis function near the origin and this will allow an order estimate to be obtained for the pointwise smoother of data functions error on a closed, bounded, infinite data region. We start with Riesz data functions:

Corollary 152 *Suppose a weight function has property W2 for some $\kappa \geq 0$ and the (real-valued) basis function G has the smoothness property assumed by Theorem 59 i.e. assume for some $s > 0$ and C_G, $h_G > 0$ that*

$$G(0) - G(x) \leq C_G |x|^{2s}, \quad |x| \leq h_G. \tag{4.19}$$

Further, suppose $\mathcal{S}_X^e$ is the Exact smoother mapping with smoothing coefficient ρ and independent data set X contained in the bounded, closed, infinite data set K.

Then if $h_{X,K} = \max\limits_{x \in K} \operatorname{dist}(x, X) < h_G$ it follows that

$$|R_y(x) - (\mathcal{S}_X^e R_y)(x)| \leq \left(\sqrt{\rho N} + k_G (h_{X,K})^s\right)^2, \quad x, y \in K, \tag{4.20}$$

where

$$k_G = (2\pi)^{-\frac{d}{4}} \sqrt{2C_G}.$$

Also

$$|R_y(x) - (\mathcal{S}_X^e R_y)(x)| \leq R_0(0), \quad x, y \in \mathbb{R}^d. \tag{4.21}$$

Proof. Let $X = \left\{x^{(k)}\right\}_{k=1}^N$ and suppose $h_{X,K} \leq h_G$. Then for each $x \in K$ there exists $x^{(k)} \in X$ such that $G(0) - G\left(x - x^{(k)}\right) \leq C_G \left|x - x^{(k)}\right|^{2s} \leq C_G (h_{X,K})^{2s}$. Thus Theorem 151 implies that for all $x \in K$

$$R_x(x) - (\mathcal{S}_X^e R_x)(x) \leq \left(\sqrt{\rho N} + \frac{\sqrt{2}}{(2\pi)^{\frac{d}{4}}} \sqrt{G(0) - G\left(x - x^{(k)}\right)}\right)^2 = \left(\sqrt{\rho N} + k_G (h_{X,K})^s\right)^2.$$

Suppose $y \in K$. Then setting $f_d = R_y$ in inequality 4.13 I get

$$\begin{aligned} \left|R_y(x) - \left(\mathcal{S}^e_X R_y\right)(x)\right| &\leq \sqrt{\left(R_y - \mathcal{S}^e_X R_y, R_y\right)_{w,0}} \sqrt{R_x(x) - \left(\mathcal{S}^e_X R_x\right)(x)} \\ &= \sqrt{R_y(y) - \left(\mathcal{S}^e_X R_y\right)(y)} \sqrt{R_x(x) - \left(\mathcal{S}^e_X R_x\right)(x)} \\ &\leq \left(\sqrt{\rho N} + k_G \left(h_{X,K}\right)^s\right)^2. \end{aligned}$$

Finally 4.21 follows from 4.14 with $f_d = R_y$. ∎

When $\rho = 0$ the above results give an interpolant error estimates of order $2s$. These are the interpolant error estimates obtained in Chapter 2. The next corollary is a simple consequence of inequality 4.13 and Corollary 152. This result is an estimate of the smoother error for an arbitrary data function in X^0_w.

Corollary 153 *Suppose:*

1. *The basis function G has the properties assumed in Corollary 152 for the constants C_G, h_G, $s > 0$.*
2. *Suppose s_e is the Exact smoother of the arbitrary data function $f_d \in X^0_w$ on the independent data set X contained in the bounded, closed, infinite data set K.*

Then if $h_{X,K} = \max\limits_{x \in K} \operatorname{dist}(x, X) \leq h_G$ we have

$$\left|f_d(x) - s_e(x)\right| \leq \sqrt{\left(f_d - s_e, f_d\right)_{w,0}} \left(\sqrt{\rho N} + k_G \left(h_{X,K}\right)^s\right), \quad x \in K, \ f_d \in X^0_w, \tag{4.22}$$

where $k_G = (2\pi)^{-\frac{d}{4}} \sqrt{2C_G}$ and $\sqrt{\left(f_d - s_e, f_d\right)_{w,0}} < \left\|f_d\right\|_{w,0}$.

Remark 154

1. *When $\rho = 0$ the last two corollaries yield the Type 1 interpolant error estimates of Subsection 2.5.2. Thus we will label the parameter s the order of convergence.*
2. *The factor $\sqrt{\left(f_d - s_e, f_d\right)_{w,0}}$ on the right side of 4.22 enables the error estimate 4.20 for R_x to be recovered immediately. This is because*

$\left(R_x - \mathcal{S}^e_X R_x, R_x\right)_{w,0} = \left(R_x - \mathcal{S}^e_X R_x\right)(x)$ *where $\mathcal{S}^e_X$ is the Exact smoother mapping.*

Error summary

Now we combine the smoother bound of Theorem 146 with the error estimates derived in this Subsection 4.6.1 to obtain the error estimates that will be tested numerically in Section 4.8.

Theorem 155 *Suppose K is a bounded, closed, infinite set and there exist constants C_G, s, $h_G > 0$ such that*

$$G(0) - G(x) \leq C_G |x|^{2s}, \quad |x| \leq h_G, \ x \in K.$$

Set $k_G = (2\pi)^{-\frac{1}{4}} \sqrt{2C_G}$. Then the Exact smoother $\mathcal{S}^e_X f_d$ satisfies the error estimates

$$\left|f_d(x) - \left(\mathcal{S}^e_X f_d\right)(x)\right| \leq \min \begin{cases} \left\|f_d\right\|_{w,0} \left(\sqrt{\rho N} + k_G \left(h_{X,K}\right)^s\right), \\ \left\|f_d\right\|_{w,0} \sqrt{R_0(0)}, \\ \left\|f_d\right\|_{\infty,K} + \left\|f_d\right\|_{w,0} \sqrt{R_0(0)} \min\left\{1, \frac{R_0(0)}{\rho}\right\}, \end{cases} \tag{4.23}$$

when $x \in K$ and $h_{X,K} = \max\limits_{s \in K} \operatorname{dist}(s, X) \leq h_G$. Here $R_0(0) = (2\pi)^{-\frac{1}{2}} G(0)$ and X is an independent data set contained in K.

We also have the corresponding double order convergence estimate

$$\left|R_y(x)-\left(\mathcal{S}_X^e R_y\right)(x)\right| \leq \min \begin{cases}\left(\sqrt{\rho N}+k_G\left(h_{X,K}\right)^s\right)^2, \\ R_0(0), \\ \left\|R_y\right\|_{\infty,K}+R_0(0) \min \left\{1, \frac{R_0(0)}{\rho}\right\},\end{cases} \tag{4.24}$$

for $x, y \in K$.

Proof. From Corollary 153 and 4.14 of Theorem 145 we have

$$\left|f_d(x)-s_e(x)\right| \leq \min \left\{\left\|f_d\right\|_{w,0}\left(\sqrt{\rho N}+k_G\left(h_{X,K}\right)^s\right), \left\|f_d\right\|_{w,0} \sqrt{R_0(0)}\right\}.$$

From Theorem 146

$$\left|s_e(x)\right| \leq \left\|f_d\right\|_{w,0} \sqrt{R_0(0)} \min \left\{1, \frac{R_0(0)}{\rho}\right\},$$

so that

$$\left|f_d(x)-s_e(x)\right| \leq \left|f_d(x)\right|+\left|s_e(x)\right| \leq \left\|f_d\right\|_{\infty,K}+\left\|f_d\right\|_{w,0} \sqrt{R_0(0)} \min \left\{1, \frac{R_0(0)}{\rho}\right\},$$

which yields 4.23. Thus when $f_d=R_y$ and $x \in K$

$$\begin{aligned}\left|R_y(x)-s_e(x)\right| &\leq \left\|R_y\right\|_{\infty,K}+\left\|R_y\right\|_{w,0} \sqrt{R_0(0)} \min \left\{1, \frac{R_0(0)}{\rho}\right\} \\ &= \left\|R_y\right\|_{\infty,K}+R_0(0) \min \left\{1, \frac{R_0(0)}{\rho}\right\},\end{aligned}$$

which combined with the estimates of Corollary 152 gives the second set of inequalities 4.24. ■

Smoother convergence

I will now introduce the concept of the *convergence of a smoother to its data function.* Theorem 73 showed that for a bounded, closed, infinite data region K there exists a nested sequence of independent data sets $X^{(k)} \subset X^{(k+1)} \subset K$ such that $h_{X^{(k)},K} \to 0$ as $k \to \infty$.

An inspection of the error estimates of this subsection clearly shows that the smoothing error can tend to zero whilst the smoothing coefficient ρ remains positive. Consider the estimate 4.22 of Corollary 153 and suppose $X^{(k)}$ is a nested sequence of independent data sets such that $h_{X^{(k)},K} \to 0$. Denote the number of points in $X^{(k)}$ by N_k. Then given $\varepsilon > 0$ there exists k_ε such that $k_G \left\|f_d\right\|_{w,0} \left(h_{X^{(k_\varepsilon)},K}\right)^s \leq \frac{\varepsilon}{2}$ when $k \geq k_\varepsilon$. Also for $N \geq N_{k_\varepsilon}$ we have $h_{X^{(k)},K} \leq h_{X^{(k_\varepsilon)},K}$. Choosing ρ_k so that $\sqrt{\rho_k N_k} = \frac{\varepsilon}{2\left\|f_d\right\|_{w,0}}$ an easy calculation shows that

$$\begin{aligned}\left|f_d(x)-\mathcal{S}_{X^{(k)}}^e f_d(x)\right| &\leq \left\|f_d\right\|_{w,0}\left(\sqrt{\rho N}+k_G\left(h_{X,K}\right)^s\right) \\ &\leq \left\|f_d\right\|_{w,0} \sqrt{\rho_k N_k}+\left\|f_d\right\|_{w,0} k_G\left(h_{X^{(k)},K}\right)^s \\ &= \varepsilon.\end{aligned}$$

We call s the *order of convergence of the smoother.*

4.6.2 Type 1 examples

The interpolant weight function examples are also used here i.e. the radial shifted thin-plate splines, Gaussian and Sobolev splines, and the tensor product extended B-spline weight functions, augmented by the tensor product central difference weight functions introduced in Chapter 3.

When $\rho = 0$ the error estimates of this subsection are the Type 1 interpolation estimates of Subsubsection 2.5.2 and consequently the values for s, C_G, h_G, k_G have already been calculated and are exhibited

Exact smoother convergence order estimates				
Type 1 non-unisolvent estimates assuming $\kappa \geq 0$: $k_G = (2\pi)^{-\frac{d}{4}} \sqrt{2C_G}$.				
Weight function	Parameter constraints	Converg. order s	C_G	h_G
Sobolev splines	$v - \frac{d}{2} = 1$	$\frac{1}{2}$	$\frac{\lVert \rho K_0(\rho) \rVert_\infty}{2^{v-1}\Gamma(v)}$ (2)	∞
$(v > d/2)$	$v - \frac{d}{2} \neq 1$	1	$\frac{\lVert D^2 \widetilde{K}_{v-d/2} \rVert_\infty}{2^v \Gamma(v)}$ (2)	"
Shifted thin-plate $(-d/2 < v < 0)$	-	1	eq. (4.25)	∞
Gaussian	-	1	$2e^{-3/2}$	∞
Extended B-spline	-	$\frac{1}{2}$	$G_1(0)^{d-1} \lVert DG_1 \rVert_\infty \sqrt{d}$ (1)	∞
Central difference	-	$\frac{1}{2}$	$(2\pi)^{-\frac{d}{4}} \sqrt{2G_1(0)^{d-1} \lVert DG_1 \rVert_\infty \sqrt[4]{d}}$ (1)	∞
(1) G_1 is the univariate basis function used to form the tensor product				
(2) K_v is the modified Bessel function and $\widetilde{K}_v(r) = r^v K_v(r)$.				

TABLE 4.1.

in Table 2.1. These values are given below in Table 4.1 augmented by the values obtained for the central difference tensor product weight functions from Table 3.1.

Equations 4.25 are referenced by Table 4.1:

$$C_G = \left|(2rf'' + f')(r_{\max})\right|, \quad \textit{where } f(r) = \left(1 + r^2\right)^v \textit{ and } r_{\max} = \frac{(1-2v)}{3}. \tag{4.25}$$

4.6.3 Type 2 error estimates ($\kappa \geq 1$)

The Corollaries 152 and 153 which were used to obtain Type 2 interpolant error estimates are again used here by Theorems 156, 157 and 158 to obtain estimates for $G(0) - G(x)$ when $\kappa \geq 1$.

Theorem 156 *Suppose a weight function satisfies property W2 for $\kappa = 1$ and denote the basis function by G. Then the smoother error estimates of Corollaries 152 and 153 hold where*

$$C_G = -\frac{1}{2}\left(\nabla^2 G\right)(0)\, d, \quad s = 1, \quad h_G = \infty, \quad k_G = (2\pi)^{-\frac{d}{4}} \sqrt{-\left(\nabla^2 G\right)(0)}\sqrt{d}. \tag{4.26}$$

Proof. Theorem 39 showed that if a weight function has $\kappa \geq 1$ then

$$G(0) - G(x) \leq -\frac{d}{2}\left(\nabla^2 G\right)(0)\, |x|^2, \quad x \in \mathbb{R}^d. \tag{4.27}$$

An application of Corollary 152 now proves this theorem directly. ∎

However, if the weight function is radial we can use the estimates:

Theorem 157 *Suppose a* ***radial weight function*** *satisfies property W2 for $\kappa = 1$ and denote the (radial) basis function by G. Set $r = |x|$. Then:*

1. *If $G(x) = f\left(r^2\right)$ then*

$$G(0) - G(x) \leq -f'(0)\, d^2 |x|^2, \quad x \in \mathbb{R}^d. \tag{4.28}$$

2. *If $G(x) = g(r)$ then*

$$G(0) - G(x) \leq -\frac{1}{2} g''(0)\, d^2 |x|^2, \quad x \in \mathbb{R}^d, \tag{4.29}$$

Further, the smoother error estimates of Corollaries 152 and 153 hold for

$$C_G = -f'(0)\, d^2 \textit{ or } -\frac{1}{2} g''(0)\, d^2, \quad s = 1, \quad h_G = \infty, \quad k_G = (2\pi)^{-\frac{d}{4}} \sqrt{2C_G}. \tag{4.30}$$

Proof. Follows directly from Theorem 39 and Corollary 152. ∎

If the weight function is a tensor product the following result will be useful:

Theorem 158 *Suppose a* ***tensor product weight function*** *satisfies property W2 for $\kappa = 1$ and denote the univariate basis function by G_1. Then*

$$G(0) - G(x) \leq -\frac{d}{2} G_1(0)^{d-1} D^2 G_1(0) |x|^2, \quad x \in \mathbb{R}^d,$$

and the smoother error estimates of Corollaries 152 and 153 hold for

$$C_G = -\frac{d}{2} G_1(0)^{d-1} D^2 G_1(0), \quad s = 1, \quad h_G = \infty, \quad k_G = (2\pi)^{-\frac{d}{4}} \sqrt{-G_1(0)^{d-1} D^2 G_1(0)} \sqrt{d}.$$

Proof. Since G_1 is the univariate basis function

$$\begin{aligned}\left(\nabla^2 G\right)(0) = \sum_{k=1}^{d} \left(D_k^2 G\right)(0) = \sum_{k=1}^{d} D_k^2 \left(G_1(x_1) G_1(x_2) \ldots G_1(x_d)\right)(0) &= \sum_{k=1}^{d} G_1(0)^{d-1} D^2 G_1(0) \\ &= G_1(0)^{d-1} D^2 G_1(0) d,\end{aligned}$$

The estimates of this theorem then follow from Theorem 156, 4.27 and an application of Corollary 152 . ■

4.6.4 Type 2 examples

When $\rho = 0$ the smoother error estimates of Corollaries 152 and 153 become algebraically identical to the Type 2 interpolant error estimates of Subsubsection 2.5.2. Further, the weight function examples used above were also used for the Type 2 interpolation examples. If we use the 'radial' Theorem 157 to do the estimates for the radial basis functions and Theorem 156 to do the estimates for the tensor product basis functions then the values of the variables s, C_G, h_G, k_G will match those obtained for the interpolants. These are given below in Table 4.2 which is a copy of Table 2.2 augmented by the results for the central difference tensor product weight functions from Table 3.1.

Exact smoother convergence order estimates			
Type 2 non-unisolvent estimates assuming $\kappa \geq 1$			
Weight function	Parameter constraints	Converg. order	$(2\pi)^{d/4} k_G / \sqrt{d}$
Sobolev splines $(v > d/2)$	$v - \frac{d}{2} \geq 2$	1	$\sqrt{\frac{\Gamma(v-d/2-1)}{2^{d/2+1}\Gamma(v)}}$
	$1 < v - \frac{d}{2} < 2$	1	$\sqrt{\frac{\Gamma(v-d/2-1)}{2^{2v-d/2-3}\Gamma(v)}}$
Shifted thin-plate $(-d/2 < v < 0)$	-	1	$\sqrt{-2v}$
Gaussian	-	1	$\sqrt{2}$
Extended B-spline $(1 \leq n \leq l)$	$n \geq 2$	1	$\sqrt{-G_1(0)^{d-1} D^2 G_1(0)}^{(1)}$
Central difference $(1 \leq n \leq l)$	$n \geq 2$	1	$\sqrt{-G_1(0)^{d-1} D^2 G_1(0)}^{(1)}$
(1) G_1 is the univariate basis function used to form the tensor product.			

TABLE 4.2.

When $\rho = 0$ we see that the order of convergence is at least 1 for an arbitrary data function and at least 2 for a Riesz representer data function, no matter what value κ takes. However, in the next section we will show that by assuming the independent data is unisolvent of order $\kappa \geq 1$ it follows that an order of convergence of at least $\lfloor \kappa \rfloor$ can be attained for an arbitrary data function and an order of convergence of $\lfloor 2\kappa \rfloor$ for a Riesz representer data function.

4.7 Pointwise error estimates obtained using unisolvent data sets

In this section we will derive smoother error results using the Lagrange interpolation theory. These results require that the independent data points constitute a unisolvent set and assumes that the boundary of the data region boundary satisfies the cone condition.

These results achieve the same order of convergence as the interpolation results of Subsection 2.5.3 in the sense that when we set $\rho = 0$ we get the same interpolation error estimates.

Unisolvent sets were introduced in Subsection 2.5.3 where they were used to define the Lagrange interpolation operators $\mathcal{P}$ and $\mathcal{Q} = I - \mathcal{P}$. Lemma 72 supplied the key interpolation theory and I reproduce it here:

Lemma 159 *(Copy of Lemma 72) Suppose first that:*

1. *Ω is a bounded, open, connected subset of $\mathbb{R}^d$ having the cone property.*
2. *X is a unisolvent subset of Ω of order m.*

Suppose $\{l_j\}_{j=1}^{M}$ is the cardinal basis of P_m with respect to a minimal unisolvent subset of Ω. Again, using Lagrange polynomial interpolation techniques, it can be shown there exists a constant $K'_{\Omega,m} > 0$ such that

$$\sum_{j=1}^{M} |l_j(x)| \leq K'_{\Omega,m} \quad , x \in \Omega,$$

and all minimal unisolvent subsets of Ω.

Now define the data point density measure

$$h_{X,\Omega} = \sup_{\omega \in \Omega} \operatorname{dist}(\omega, X),$$

and fix $x \in X$. By using Lagrange interpolation techniques it can be shown there are constants $c_{\Omega,m}, h_{\Omega,m} > 0$ such that when $h_{X,\Omega} < h_{\Omega,m}$ there exists a minimal unisolvent set $A \subset X$ satisfying

$$\operatorname{diam}(A \cup \{x\}) \leq c_{\Omega,m} h_{X,\Omega}.$$

Our main result for arbitrary data functions is:

Theorem 160

1. *Let w be a weight function with property W2 for some parameter $\kappa \geq 1$ and let G be the basis function generated by w. Set $m = \lfloor \kappa \rfloor$.*
2. *Suppose s_e is the Exact smoother of an arbitrary data function $f_d \in X_w^0$ evaluated on the independent data set X contained in the data region Ω.*
3. *Suppose the notation and assumptions of Lemma 159 hold for this theorem i.e. X is m-unisolvent and Ω is a bounded, open, connected set whose boundary satisfies the cone condition.*

Then there exist positive constants $k_G, c_{\Omega,m}, K'_{\Omega,m}, h_{\Omega,m}$ such that the smoothing error satisfies

$$|f_d(x) - s_e(x)| < \sqrt{(f_d - s_e, f_d)_{w,0}} \left(K'_{\Omega,m} \sqrt{\rho N} + k_G (h_{X,\Omega})^m \right), \quad x \in \overline{\Omega}, \tag{4.31}$$

when $h_{X,\Omega} = \sup_{\omega \in \Omega} \operatorname{dist}(\omega, X) < h_{\Omega,m}$.

Here $k_G = \frac{d^{m/2}}{(2\pi)^{d/2}} (c_{\Omega,m})^m K'_{\Omega,m} \max_{|\beta|=m} \left| D^{2\beta} G(0) \right|$ and the constants $c_{\Omega,m}, K'_{\Omega,m}$ and $h_{\Omega,m}$ only depend on Ω, m and d. In terms of the integrals which define weight property W2 I have

$$\max_{|\beta|=m} \left| D^{2\beta} G(0) \right| \leq (2\pi)^{-d/2} \int \frac{|\xi|^{2m} d\xi}{w(\xi)}.$$

Proof. Fix $x \in \Omega$. By Lemma 159 there exists a minimal unisolvent set $A = \{a_k\}_{k=1}^{M}$ of order m and constants $c_{\Omega,m}, h_{X,\Omega} > 0$ such that

$$\operatorname{diam}(A \cup \{x\}) \leq c_{\Omega,m} h_{X,\Omega} \quad \textit{when } h_{X,\Omega} < h_{\Omega,m}.$$

Further, there exists a positive constant $K'_{\Omega,m}$ such that

$$\sum_{j=1}^{M} |l_j(x)| \leq K'_{\Omega,m}, \quad x \in \Omega,$$

where $\{l_j\}_{j=1}^{M}$ is the cardinal basis of P_m with respect to A. The next step is to partition the error function using the Lagrange interpolation projections $\mathcal{P}$ and $\mathcal{Q} = I - \mathcal{P}$ so that

$$f_d - s_e = \mathcal{P}(f_d - s_e) + \mathcal{Q}(f_d - s_e),$$

and note that $\mathcal{P}(f_d - s_e)$ is not zero. However, I can still apply Lemma 159 to the second term $\mathcal{Q}(f_d - s_e)$ in exactly the the same way as in the proof of Theorem 74 to show that

$$|\mathcal{Q}(f_d - s_e)(x)| \leq k_G \|f_d - s_e\|_{w,0} (h_{X,\Omega})^m, \tag{4.32}$$

when $h_{X,\Omega} < h_{\Omega,m}$. Part 2 of Lemma 142 then implies $\|s_e\|_{w,0}^2 \leq (s_e, f_d)_{w,0}$ and so

$$|\mathcal{Q}(f_d - s_e)(x)| \leq k_G (f_d, f_d - s_e)_{w,0} (h_{X,\Omega})^m,$$

when $h_{X,\Omega} < h_{\Omega,m}$. However, it still remains to estimate $\mathcal{P}(f_d - s_e)(x)$. By definition of $\mathcal{P}$

$$\mathcal{P}(f_d - s_e)(x) = \sum_{k=1}^{M} (f_d(a_k) - s_e(a_k)) l_k(x),$$

and by 4.13

$$|f_d(a_k) - s_e(a_k)| \leq \sqrt{(f_d - s_e, f_d)_{w,0}} \sqrt{R_{a_k}(a_k) - (\mathcal{S}_X^e R_{a_k})(a_k)}. \tag{4.33}$$

But by Lemma 159, $\sum_{k=1}^{M} |l_k(x)| \leq K'_{\Omega,m}$ for $x \in \Omega$, so

$$\begin{aligned}
|\mathcal{P}(f_d - \mathcal{S}_X^e f_d)(x)| &\leq \sum_{k=1}^{M} |f_d(a_k) - s_e(a_k)| \, |l_k(x)| \\
&\leq \left(\sum_{k=1}^{M} |l_k(x)| \right) \max_{k=1}^{M} |f_d(a_k) - s_e(a_k)| \\
&\leq K'_{\Omega,m} \sqrt{(f_d, f_d - s_e)_{w,0}} \ \max_{k=1}^{M} \sqrt{(R_{a_k}(a_k) - (\mathcal{S}_X^e R_{a_k})(a_k))},
\end{aligned}$$

where $\mathcal{S}_X^e$ is Exact smoother operator. Hence Lemma 149 implies

$$|\mathcal{P}(f_d - \mathcal{S}_X^e f_d)(x)| \leq K'_{\Omega,\kappa} \sqrt{(f_d - s_e, f_d)_{w,0}} \sqrt{\rho N}. \tag{4.34}$$

Combining inequalities 4.32 and 4.34 gives inequality 4.31 for all $x \in \Omega$. The extension of the inequality to $\overline{\Omega}$ is an easy consequence of the fact that f_d and s_e are continuous on $\mathbb{R}^d$. ∎

Observe that when $\rho = 0$ the order of convergence derived above is $\lfloor \kappa \rfloor$ and this is the same order derived for the interpolant of an arbitrary data function in Theorem 74. I now derive a double order of convergence result for the Riesz representer data functions R_y. Note that the factor $\sqrt{(f_d - \mathcal{S}_X^e f_d, f_d)_{w,0}}$ in the right side of the estimate 4.31 for $|f_d(x) - (\mathcal{S}_X^e f_d)(x)|$ facilitates the following simple proof:

Corollary 161 *Under the conditions and notation of Theorem 160 the estimate*

$$R_y(x) - (\mathcal{S}_X^e R_y)(x) \leq \left(K'_{\Omega,m} \sqrt{\rho N} + k_G (h_{X,\Omega})^m \right)^2, \quad x, y \in \overline{\Omega},$$

holds when $h_{X,\Omega} < h_{\Omega,m}$ and $m = \lfloor \kappa \rfloor$.

Proof. Suppose $h_{X,\Omega} < h_{\Omega,m}$. When $f_d = R_x$ the inequality proved in Theorem 160 becomes

$$\left(R_x - \mathcal{S}_X^e R_x\right)(x) \leq \sqrt{\left(R_x - \mathcal{S}_X^e R_x\right)(x)} \left(K'_{\Omega,m}\sqrt{\rho N} + k_G \left(h_{X,\Omega}\right)^m\right), \quad x \in \overline{\Omega},$$

which implies that

$$\left(R_x - \mathcal{S}_X^e R_x\right)(x) \leq \left(K'_{\Omega,m}\sqrt{\rho N} + k_G \left(h_{X,\Omega}\right)^m\right)^2,$$

and an application of the estimate 4.13 with $f_d = R_y$ gives the desired result. ∎

4.7.1 Examples

The weight function examples used here are those used in for interpolation i.e. the radial shifted thin-plate splines, Gaussian and Sobolev splines, and the tensor product extended B-spline weight functions, augmented by the tensor product central difference weight functions from Chapter 3.

When when the smoothing coefficient ρ is zero the error estimates of this section are the unisolvent interpolation estimates of Subsection 2.5.3 and consequently values for κ and the Lagrangian constants $k_G, c_{\Omega,m}, K'_{\Omega,m}, h_{\Omega,m}$ are the same. We can't calculate the Lagrangian constants but we estimated the convergence order for the interpolant as $\max\lfloor\kappa\rfloor$ which is given in Table 2.3. We use this value here and they are given below in Table 4.3 augmented by the order values obtained for the central difference tensor product weight functions from Table 3.1.

Exact smoother convergence order estimates			
Unisolvent data and $\kappa \geq 1$			
Weight function	Parameter constraints	Convergence order $(\max\lfloor\kappa\rfloor)$	$h_{\Omega,\max\lfloor\kappa\rfloor}$
Sobolev splines $(v > d/2)$	$v - \frac{d}{2} = 2, 3, 4, \ldots$	$\lfloor v - \frac{d}{2} \rfloor - 1$	∞
	$v - \frac{d}{2} > 1,\ v - \frac{d}{2} \notin \mathbb{Z}_+$	$\lfloor v - \frac{d}{2} \rfloor$	"
Shifted thin-plate	-	$2, 3, 4, \ldots$	∞
Gaussian	-	$2, 3, 4, \ldots$	∞
Extended B-spline $(1 \leq n \leq l)$	$n \geq 2$	$n - 1$	∞
Central difference $(1 \leq n \leq l)$	$n \geq 2$	$n - 1$	∞

TABLE 4.3.

4.8 Numerical experiments with the extended B-splines (non-unisolvent case)

In this section the convergence of the Exact smoother to its data function is studied numerically using scaled, extended B-splines, an example of which is the hat function. We will use the same 1-dimensional data functions and the same extended B-splines, introduced in Theorem 7, that were used for the numerical interpolation experiments described in Section 2.6.

We will only consider the numerical experiments in one dimension so that our standard data density parameter $h_{X,K} = \max\limits_{x \in K} \operatorname{dist}(x, X)$ can be easily calculated. It is also easier to test the non-unisolvent error estimates i.e. those derived without assuming unisolvent data sets. This is because in the non-unisolvent case the constants are known precisely. In the case of unisolvent data the theory is complex and it is unclear what are suitable upper bounds for the constants e.g. $K'_{\Omega,m}$, $c_{\Omega,m}$, $h_{\Omega,m}$ in Theorem 160.

Recall from Theorem 7 that the extended B-splines are a special class of tensor product weight functions which satisfy property W2. Indeed, for given integers l, n their univariate weight function is defined by

$$w_1(x) = \prod_{i=1}^{d} \frac{x_i^{2n}}{\sin^{2l} x_i}, \quad x = (x_1, \ldots, x_d) \in \mathbb{R}^d, \tag{4.35}$$

where the weight function w has property W2 for κ iff n and l satisfy

$$\kappa + 1/2 < n \leq l. \tag{4.36}$$

Theorem 29 showed that the basis function G generated by w is the tensor product $G(x) = \prod_{k=1}^{d} G_1(x_k)$ where

$$G_1(t) = (-1)^{l-n} \tfrac{(2\pi)^{l/2}}{2^{2(l-n)+1}} D^{2(l-n)} \left((*\Lambda)^l \right) \left(\tfrac{t}{2} \right), \quad t \in \mathbb{R}^1, \tag{4.37}$$

and $(*\Lambda)^l$ denotes the convolution of l 1-dimensional hat functions. Further, $G_1 \in C_0^{(2n-2)}\left(\mathbb{R}^1\right)$ and $D^{2n-1}G_1$ is a piecewise constant function. Finally, $G \in C_0^{(2n-2)}\left(\mathbb{R}^d\right)$ and the derivatives $\{D^\alpha G\}_{|\alpha|=2n-1}$ are bounded functions.

Because all the extended B-spline basis weight functions 4.35 have a power of $\sin x$ in the denominator we will need the special classes of data functions developed in Section 2.6.

The 1-dimensional independent X data set is constructed using a uniform distribution on the interval $K = [-1.5, 1.5]$. Each of 20 data files is exponentially sampled using a multiplier of approximately 1.2 and a maximum of 5000 points, and then we plot $\log_{10} h_{X,K}$ against $\log_{10} N$ where $N = |X|$. It then seems quite reasonable to use a least-squares linear fit and in this case we obtain

$$h_{X,K} \simeq 3.09 N^{-0.81}. \tag{4.38}$$

For ease of calculation let

$$h_{X,K} = h_1 N^{-a}, \quad h_1 = 3.09, \quad a = 0.81. \tag{4.39}$$

Noting the error estimates of Theorem 155 we use 4.39 to write $h_{X,K}^s = h_1^s N^{-as}$ and define the Exact smoother error estimates

$$|f_d(x) - (\mathcal{S}_X^e f_d)(x)| \lesssim E_N(\rho), \quad x \in K, \tag{4.40}$$

and

$$|R_y(x) - (\mathcal{S}_X^e R_y)(x)| \lesssim \varepsilon_N(\rho), \quad x, y \in K, \tag{4.41}$$

where

$$E_N(\rho) = \min \begin{cases} \|f_d\|_{w,0} \left(\sqrt{\rho N} + k_G h_1^s N^{-as} \right), \\ \|f_d\|_{w,0} \sqrt{R_0(0)}, \\ \|f_d\|_{\infty,K} + \|f_d\|_{w,0} \sqrt{R_0(0)} \min\left\{1, \frac{R_0(0)}{\rho}\right\}, \end{cases} \tag{4.42}$$

and

$$\varepsilon_N(\rho) = \min \begin{cases} \left(\sqrt{\rho N} + k_G h_1^s N^{-as} \right)^2, \\ R_0(0), \\ \|R_y\|_{\infty,K} + R_0(0) \min\left\{1, \frac{R_0(0)}{\rho}\right\}. \end{cases} \tag{4.43}$$

4.8.1 Extended B-splines with $n = 1$

The case $n = 1$, $l = 1$

The 1-dimensional hat weight function is given by $w_\Lambda(\xi) = \sqrt{2\pi} \left(\frac{\xi/2}{\sin(\xi/2)} \right)^2$ (1.5) and examining 4.37 we see that it is a scaled, extended B-spline weight function corresponding to the parameters $n = l = 1$. Also 4.36 implies $\max \lfloor \kappa \rfloor = 0$.

Now suppose Π is the 1-dimensional rectangular function given by $\Pi(x) = 1$ when $|x| < 1/2$ and $\Pi(x) = 0$ when $|x| > 1/2$. Then it was shown in Subsubsection 2.6.1 that:

Theorem 162 ***Data function*** *Suppose w_Λ is the univariate hat weight function. Then $u * \Pi \in X^0_{w_\Lambda}$ if $u \in L^2(\mathbb{R}^1)$. Further, we can define $V \in L^1_{loc} \cap C^{(0)}_{BP}$ by*

$$V(x) = \int_0^x u(t)\,dt, \quad u \in L^2,$$

so that $DV = u$ a.e. and

$$u * \Pi = (2\pi)^{-\frac{1}{2}} \left(V\left(x + \frac{1}{2}\right) - V\left(x - \frac{1}{2}\right) \right).$$

To obtain our data function $f_d = u * \Pi$ we will follow the interpolation approach of Subsubsection 2.6.1 and choose

$$u = e^{-x^2},$$

for which

$$V = (2\pi)^{\frac{1}{2}} \operatorname{erf}, \qquad \|u\|_2 = 2(2\pi)^{\frac{1}{4}}, \qquad \|u * \Pi\|_{w_\Lambda,0} = 2, \tag{4.44}$$

and

$$f_d = u * \Pi = \operatorname{erf}\left(x + \frac{1}{2}\right) - \operatorname{erf}\left(x - \frac{1}{2}\right). \tag{4.45}$$

Since we are using the hat basis function, $G(0) = 1$, $C_G = 1$, $s = 1/2$, $h_G = \infty$, and from 4.44 and 4.45, $\|f_d\|_{w,0} = 2$.

For the *double rate* convergence experiment we will use $R_0 = (2\pi)^{-\frac{1}{2}} \Lambda$ as the data function.

<u>Numerical results</u>

We start by plotting the error bounds given by 4.23 and 4.24, together with the actual absolute smoother error, against the smoothing parameter. The results for the data function f_d are shown in Figure 4.1 and the results for the data function R_0 are shown in Figure 4.2:

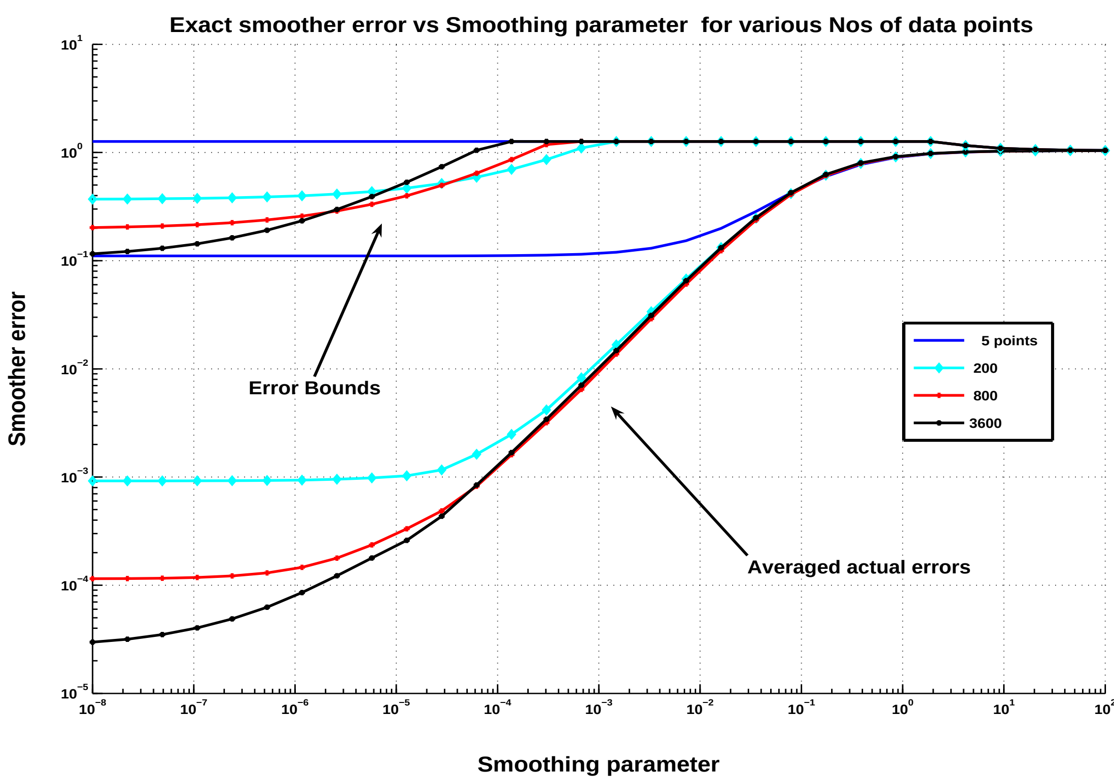

FIGURE 4.1. Exact smoother errors vs smooth parm: data func is $\operatorname{erf}\left(x + \frac{1}{2}\right) - \operatorname{erf}\left(x - \frac{1}{2}\right)$.

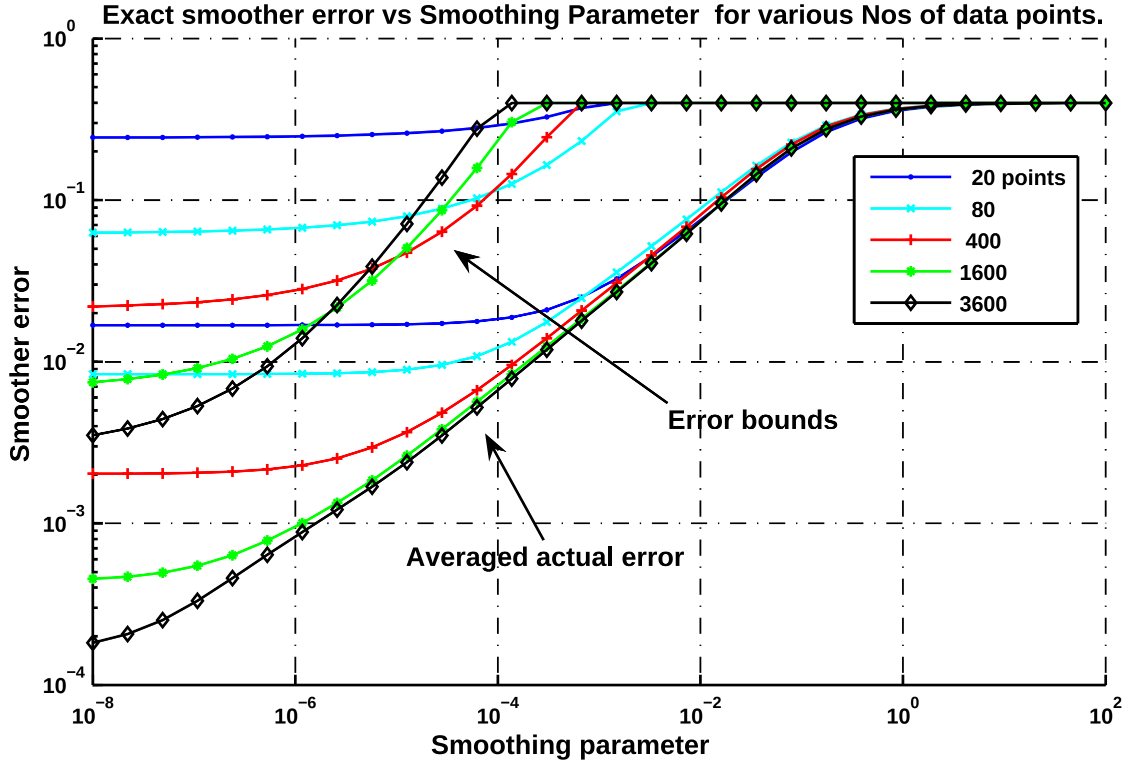

FIGURE 4.2. Exact smoother errors vs smooth param: hat basis func, data func is R_0.

The Figure 4.1 seems typical for C^{∞} data functions and it is clear that except for smoothing parameters larger than 10 the predicted error bounds given in 4.42 and 4.43 capture only a small part of the actual convergence rate. Indeed, the error ratio for the function R_0 is $10^{1.5}$ and that for f_d ranges from $10^{1.5}$ to 2×10^3.

We now move on to examine the relationship between the smoother error and the data density. To do this we need to fix a value for the smoothing parameter and we choose $\rho = 10^{-6}$. We select this value because often the Exact smoother error curve has a minimum near this value. Using the functions and parameters discussed at the end of the last subsection we obtain the four subplots displayed in Figure 4.3, each display being the superposition of 20 smoothers.

The two upper subplots relate to the data function 4.45 and the lower subplot relates to the Riesz representer data function $R_0 = (2\pi)^{-d/2} G(x - \cdot)$ (see 1.37). The right-hand subplots are filtered versions of the actual error. The data function is given at the top of the left-hand plots and the annotation at the bottom of the figure supplies the following additional information:

Input parameters

N = L = 1 - the hat function is a member of the family of scaled, extended B-splines with the indicated parameter values.

spl scale 1/2 - the actual scaling of the spline basis function is 1/2 divided by **spl scale**.

sm parm 1e-6 - the smoothing parameter is 10^{-6}.

samp 20 - the sample size. This is the number of test data files generated. The data function is evaluated on the interval $[-1.5, 1.5]$ at points selected using a uniform (statistical) distribution.

pts 2:5K - the smallest number of data points is 2 and the largest number of data points is 5000. The other values are given in exponential steps with a multiplier of approximately 1.3.

Output parameters/messages

No ill-condit - this indicates all Exact smoother matrices were always properly conditioned.

Note that all the plots shown below have the same format and annotations.

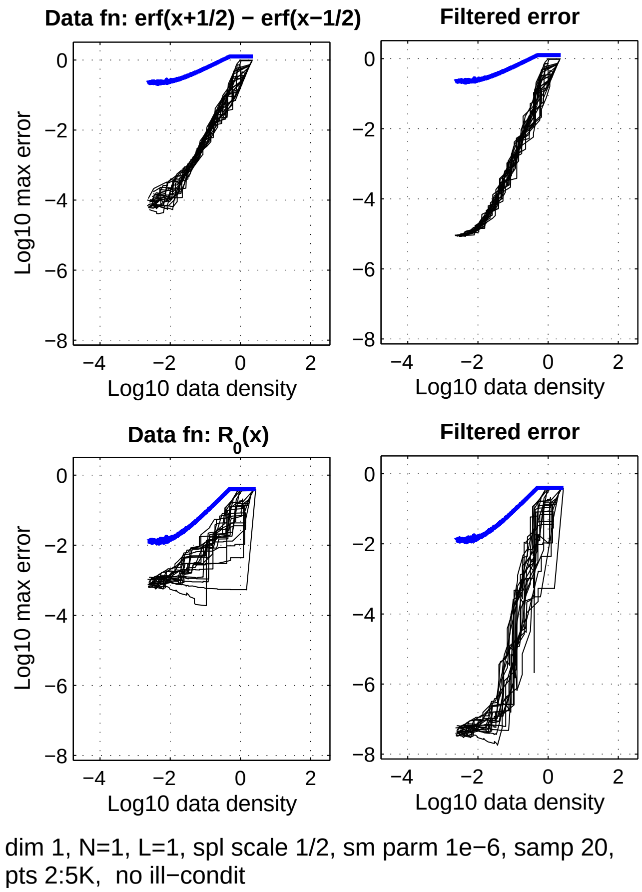

FIGURE 4.3. Convergence of Exact smoother.

As mentioned above the smoother is filtered. The filter calculates the value below which 80% of the errors lie. The filter is designed to remove 'large', isolated spikes which dominate the actual smoother errors. The smoother error was calculated on a grid with 300 cells applied to the domain of the data function. No filter is used for the first five smoothers because there is no instability for small numbers of data points. I conclude that in this case the 'spike' filter can only meaningfully applied to the smoother of R_0. This filter will be used in all the cases below.

As the number of points increases the numerical smoother of R_0 is observed to simplify to three very large increasingly narrow spikes at 0 and ± 1 and these dominate by about four orders of magnitude a stable residual error function of uniform amplitude and zero trend. The smoother of f_d consists of intermingled spikes of various heights superimposed on a trend curve of amplitude comparable to the average residual spike size. The maximum spike height is at most one order of magnitude of the average spike height so the smoother is stable.

The (blue) points above each smoother in Figure 4.3 represent the theoretical upper bound for the error given by the estimates 4.23 or 4.24, and the adjacent (red) line has a slope which gives the actual rate of convergence, ignoring instability. Clearly for the data function f_d the theoretical error bound substantially underestimates the actual error. Regarding the data function R_0, the estimated 'double' convergence rate is $2s = 1$ and the estimated (unfiltered) rate is also 1 so for $\rho = 10^{-6}$ the theoretical upper error bound for the data function R_0 is able to take into account quite closely the instability of the smoother. However, there is a large difference when the filtered error is considered.

The case $n = 1$, $l = 2$

Amongst other things the following result, extracted from Theorem 77, will allow us to generate data functions for 1-dimensional extended B-spline spaces for which the X_w^0 norm can be calculated This result is closely related to the calculations done above for the hat function.

Theorem 163 *(Portion of Theorem 77) Suppose w is the extended B-spline weight function with parameters n and l given by 1.20, and that $U \in L^2\left(\mathbb{R}^1\right)$, $D^n U \in L^2$ in the sense of distributions. Then if we define the distribution*

$$f_d = \delta_2^l U, \quad l = 1, 2, 3, \ldots,$$

where δ_2 is the central difference operator

$$\delta_2 U = U\left(\cdot + 1\right) - U\left(\cdot - 1\right), \tag{4.46}$$

it follows that $f_d \in X_w^0$ and

$$\left\|f_d\right\|_{w,0} = 2^l \left\|D^n U\right\|_2. \tag{4.47}$$

Our basis function is the extended B-spline G_1 with parameters $n = 1$ and $l = 2$ given by 4.37 i.e.

$$G_1\left(t\right) = \left(-1\right)^{l-n} \tfrac{(2\pi)^{l/2}}{2^{2(l-n)+1}} \left(D^{2(l-n)}\left(\left(*\Lambda\right)^l\right)\right)\left(\tfrac{t}{2}\right) = -\tfrac{\pi}{4}\left(D^2\left(\Lambda * \Lambda\right)\right)\left(\tfrac{t}{2}\right), \quad t \in \mathbb{R}^1.$$

But

$$\begin{aligned} D^2\left(\Lambda * \Lambda\right) &= \Lambda * D^2\Lambda = \Lambda * \left(\delta\left(\cdot + 1\right) - 2\delta + \delta\left(\cdot - 1\right)\right) \\ &= \frac{1}{\sqrt{2\pi}}\left(\Lambda\left(\cdot + 1\right) - 2\Lambda + \Lambda\left(\cdot - 1\right)\right), \end{aligned}$$

so that

$$G_1\left(t\right) = -\frac{\sqrt{2\pi}}{8}\left(\Lambda\left(\tfrac{t}{2} + 1\right) - 2\Lambda\left(\tfrac{t}{2}\right) + \Lambda\left(\tfrac{t}{2} - 1\right)\right),$$

and hence

$$DG_1\left(t\right) = -\frac{\sqrt{2\pi}}{16}\left(\Lambda'\left(\tfrac{t}{2} + 1\right) - 2\Lambda'\left(\tfrac{t}{2}\right) + \Lambda'\left(\tfrac{t}{2} - 1\right)\right),$$

i.e. $\left\|DG_1\right\|_\infty = \frac{3}{16}\sqrt{2\pi}$. Since the (distributional) derivative is bounded the distributional Taylor series expansion of Lemma 42 can be used to write

$$G_1\left(0\right) - G_1\left(t\right) \le \left\|DG_1\right\|_\infty \left|t\right|, \quad x \in \mathbb{R}^1,$$

which means that

$$G_1\left(0\right) = \frac{\sqrt{2\pi}}{4}, \; C_G = \left\|DG_1\right\|_\infty = \frac{3}{16}\sqrt{2\pi}, \; s = \frac{1}{2}, \; h_G = \infty.$$

With reference to the last theorem we will choose the bell-shaped data function

$$f_d = \delta_2^2 U \in X_w^0, \tag{4.48}$$

where

$$U\left(x\right) = \frac{e^{-k_{1,2}x^2}}{\delta_2^2\left(e^{-k_{1,2}x^2}\right)\left(0\right)} = \frac{e^{-k_{1,2}x^2}}{2\left(1 - e^{-4k_{1,2}}\right)}, \quad k_{1,2} = 0.3,$$

and

$$\left\|f_d\right\|_{w,0} = 4\left\|DU\right\|_2 = \sqrt[4]{2\pi}\frac{\sqrt[4]{4k_{1,2}}}{1 - e^{-4k_{1,2}}}.$$

The error estimates are given by 4.23 or 4.24 where $k_G = \left(2\pi\right)^{-\frac{1}{4}}\sqrt{2C_G}$ and $R_0\left(0\right) = \left(2\pi\right)^{-\frac{1}{2}} G_1\left(0\right)$.

For the theory developed in the previous sections it was convenient to use the simple, unscaled weight function definition 4.35 but for computations it may be easier to use the unscaled version of the extended B-spline basis function given in the next theorem.

Theorem 164 *Suppose $G = \left(-1\right)^{l-n} D^{2(l-n)}\left(\left(*\Lambda\right)^l\right)$ where Λ is the 1-dimensional hat function and n, l are integers such that $1 \le n \le l$.*

Then for given $\lambda > 0$, $G\left(\lambda x\right)$ is called a scaled extended B-spline basis function. The corresponding weight function is $w_\lambda\left(t\right) = 2\lambda a w\left(\frac{t}{2\lambda}\right)$ where w is the extended B-spline weight function with parameters n, l and $a = \frac{(2\pi)^{l/2}}{2^{2(l-n)+1}}$. Indeed, for w_λ we can choose $\kappa = n - 1$.

Further

$$G\left(0\right) - G\left(\lambda x\right) \le \lambda\left\|DG\right\|_\infty \left|x\right|, \quad x \in \mathbb{R}^1. \tag{4.49}$$

Finally, if $f_d \in X_w^0$ and $g_d\left(x\right) = f_d\left(2\lambda x\right)$, it follows that $g_d \in X_{w_\lambda}^0$ and $\left\|g_d\right\|_{w_\lambda,0} = \sqrt{a}\left\|f_d\right\|_{w,0}$.

Numerical results

As with the previous case we select $\rho = 10^{-6}$ and, using the spike filter of the previous case and the functions and parameters discussed in the last subsection, we obtain the four subplots displayed in Figure 4.4, each display being the superposition of 20 smoothers.

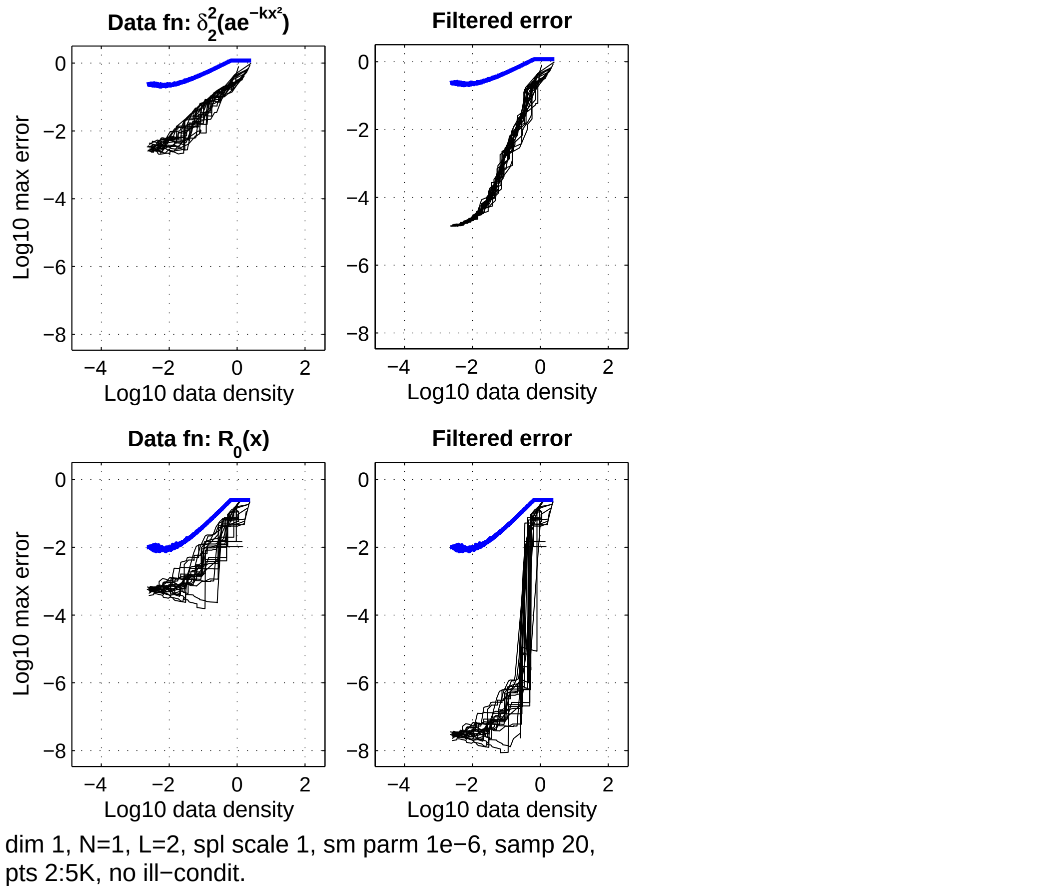

FIGURE 4.4. Convergence of Exact smoother.

These results are significantly different to those obtained for the previous case $n = l = 1$. As the number of points increases the numerical smoother of R_0 is observed to quickly simplify to a single very large increasingly narrow spike at the origin, and this dominates a stable error function of uniform amplitude and zero trend. The smoother of f_d typically forms two major spikes at the boundary $x = \pm 3/2$ by 100 data points and then two minor spikes at $x = \pm 1/2$ at about 1000 data points. These spikes are superimposed on a trend curve of amplitude comparable to the average spike size of a small stable error function.

The (blue) points above the actual errors in Figure 4.4 are the estimated upper bounds for the error given by the estimates 4.23 or 4.24, and the adjacent (red) line gives the average convergence rate. Clearly for the data function f_d the theoretical error bound substantially underestimates the convergence rate. Regarding the data function R_0, for $\rho = 10^{-6}$ the theoretical upper error bound for the data function R_0 is able to take into account quite closely the instability of the smoother. On the other hand, the filtered curves for the data function R_0 are obtained from the corresponding unfiltered curves by inserting a large step when the data density is about 0.5 and so the theoretical error estimates are at least 4 orders of magnitude greater than the actual errors.

4.8.2 Extended B-splines with $n = 2$

Since $n \geq 2$, we can use the error estimates of Theorem 158 as well as the same scaled B-spline and data function that we used for the interpolant in Subsection 4.8.2.

The case $n = 2$, $l = 2$

The basis function is

$$G_{2,2}(x) = \frac{(\Lambda * \Lambda)(2x)}{(\Lambda * \Lambda)(0)} = \frac{3\sqrt{2\pi}}{2}(\Lambda * \Lambda)(2x),$$

with scaling factor $\lambda = 4$ and is such that $\operatorname{supp} G_{2,2} = [-1, 1]$ and $G_{2,2}(0) = 1$. To calculate $G_{2,2}$ we use the formula

$$G_{2,2}(x) = (1+x)^2 \Lambda(2x+1) + \left(1-2x^2\right)\Lambda(2x) + (1-x)^2 \Lambda(2x-1),$$

and choose the bell-shaped data function

$$f_d = \delta_2^2 U \in X_w^0, \tag{4.50}$$

where

$$U(x) = \frac{e^{-k_{1,2}x^2}}{\delta_2^2\left(e^{-k_{1,2}x^2}\right)(0)} = \frac{e^{-k_{1,2}x^2}}{2\left(1-e^{-4k_{1,2}}\right)}, \quad k_{1,2} = 0.3, \tag{4.51}$$

so that

$$k_G = \frac{\sqrt{12}}{(2\pi)^{1/4}}; \quad \|f_d\|_{w,0} = \sqrt[4]{72\pi}\frac{(k_{1,2})^{3/4}}{1-e^{-4k_{1,2}}}, \quad k_{1,2} = 0.3; \quad h_G = \infty. \tag{4.52}$$

For double rate convergence experiments $R_0 = (2\pi)^{-\frac{1}{2}} G_{2,2}$ is the data function and $R_0(0) = (2\pi)^{-\frac{1}{2}}$.

The next step is to calculate $\|f_d\|_{\infty,K}$ and $\|R_0\|_{\infty,K}$ where $K = [-1.5, 1.5]$. Clearly $\|R_0\|_{\infty,K} = (2\pi)^{-\frac{1}{2}} G_{2,2}(0) = (2\pi)^{-\frac{1}{2}}$ and from 4.50 and 4.51, $\|f_d\|_{\infty,K} = f_d(0) = 1$:

$$\|R_0\|_{\infty,K} = R_0(0) = (2\pi)^{-\frac{1}{2}}, \quad \|f_d\|_{\infty,K} = 1. \tag{4.53}$$

Numerical results

Using these functions and parameters the four subplots of Figure 4.5 each display the superposition of 20 interpolants or 20 filtered interpolants as well the curves which show the upper bounds for the error given by the estimates 4.23 and 4.24.

As the number of points increases the numerical smoother of R_0 is observed to simplify to a very stable, smooth, symmetrical pattern with a single spike at the origin and two smaller spikes at $\pm 1/2$, each about $1/3$ the central spike's height. At $\pm 1$and $\pm 3/2$ there are small rounded 'spikes'. The smoother between the spikes is very smooth and the trend is zero.

The smoother of f_d forms two spikes at the boundary $x = \pm 3/2$. These spikes are superimposed on a trend curve of amplitude which dominates the amplitude of a stable residual error function. However the trend curve has a larger amplitude than the boundary spikes.

The (blue) points above each smoother in Figure 4.5 are the estimated upper bounds for the error given by 4.23 or 4.24. Clearly, in both the filtered and unfiltered cases, the theoretical bound substantially underestimates the convergence rate.

4.9 Approximation of the Exact smoother

We shall finish this chapter by approximating the Exact smoother by an easily calculated member of $W_{G,X}$ which does not involve calculating the inverse of a matrix. This result will be used to motivate the derivation of the Approximate smoother in Subsection 5.3.1 of the next chapter.

The relevant properties of the vector-valued evaluation operator $\widetilde{\mathcal{E}}_X$ and its adjoint $\widetilde{\mathcal{E}}_X^*$ w.r.t. X_w^0 were proven in Theorem 52. Now suppose $X = \left\{x^{(k)}\right\}_{k=1}^N$ is an ordered set of points in $\mathbb{R}^d$ and $f \in X_w^0$. Define $g \in X_w^0$ by

$$f = \frac{1}{NR_0(0)}\widetilde{\mathcal{E}}_X^*\widetilde{\mathcal{E}}_X f + g = \frac{1}{NR_0(0)}\sum_k f\left(x^{(k)}\right) R_{x^{(k)}} + g, \tag{4.54}$$

so that

$$f - g = \frac{1}{NR_0(0)}\widetilde{\mathcal{E}}_X^*\widetilde{\mathcal{E}}_X f, \tag{4.55}$$

has the following properties:

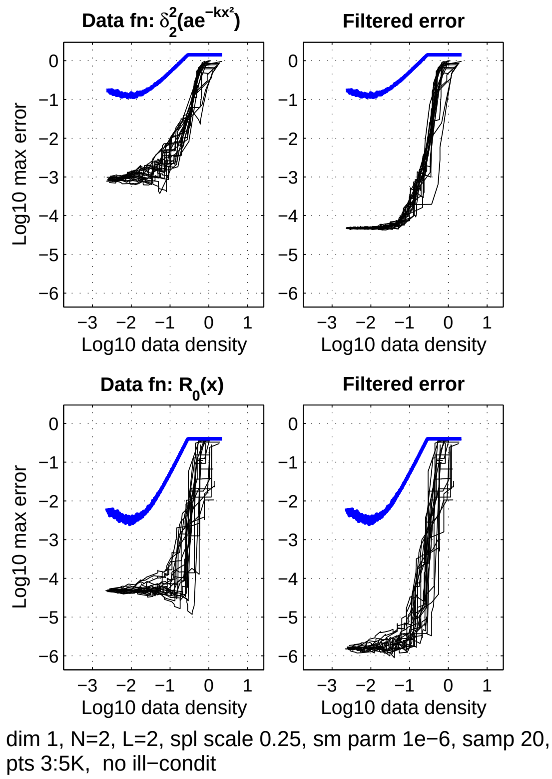

FIGURE 4.5. Convergence of the Exact smoother.

Lemma 165 *Suppose f, g satisfy equation 4.54. Then:*

1. $\|f-g\|_{w,0} \leq \|f\|_{w,0}$.

2. $|f(x)-g(x)| \leq \sqrt{R_0(0)}\, \|f\|_{w,0}$.

3. $\left|\widetilde{\mathcal{E}}_X f-\widetilde{\mathcal{E}}_X g\right| \leq \frac{1}{\sqrt{N R_0(0)}}\left|\widetilde{\mathcal{E}}_X f\right| \leq \|f\|_{w,0}$.

4. *The Exact smoother operator $\mathcal{S}_X^e$ satisfies:* $\left|\mathcal{S}_X^e f(x)-\mathcal{S}_X^e g(x)\right| \leq \sqrt{R_0(0)}\, \|f\|_{w,0}$.

Proof. By definition of the Riesz representer

$$\|R_x\|_{w,0}=\|R_0\|_{w,0}=\sqrt{R_0(0)} \tag{4.56}$$

so part 3 of Theorem 52 implies

$$\left\|\widetilde{\mathcal{E}}_X^*\right\|=\left\|\widetilde{\mathcal{E}}_X\right\|=\|R_{X,X}\| \leq \sqrt{N}\sqrt{R_0(0)}. \tag{4.57}$$

Thus:

Part 1.

$$\|f-g\|_{w,0}=\frac{1}{N R_0(0)}\left\|\widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X f\right\|_{w,0} \leq \|f\|_{w,0}.$$

Part 2.

$$\begin{aligned}|f(x)-g(x)|=\frac{1}{N R_0(0)}\left|\left(\widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X f, R_x\right)_{w,0}\right| &=\frac{1}{N R_0(0)}\left\|\widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X f\right\|_{w,0}\|R_x\|_{w,0} \\ &\leq \sqrt{R_0(0)}\, \|f\|_{w,0}.\end{aligned}$$

Part 3. By part 5 of Theorem 52, $\widetilde{\mathcal{E}}_X \widetilde{\mathcal{E}}_X^* = R_{X,X}$ so

$$\left|\widetilde{\mathcal{E}}_X f - \widetilde{\mathcal{E}}_X g\right| = \frac{1}{NR_0(0)} \left|R_{X,X} \widetilde{\mathcal{E}}_X f\right| \leq \frac{1}{NR_0(0)} \|R_{X,X}\| \left|\widetilde{\mathcal{E}}_X f\right| = \frac{1}{\sqrt{NR_0(0)}} \left|\widetilde{\mathcal{E}}_X f\right| \\ \leq \|f\|_{w,0}.$$

Part 4. From part 2 of Theorem 140, the Exact smoother operator satisfies, $\|\mathcal{S}_X^e\| \leq 1$. Hence

$$\begin{aligned}|\mathcal{S}_X^e f(x) - \mathcal{S}_X^e g(x)| = \frac{1}{NR_0(0)} \left|\left(\mathcal{S}_X^e \widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X f, R_x\right)\right| &\leq \frac{1}{NR_0(0)} \left\|\mathcal{S}_X^e \widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X f\right\|_{w,0} \|R_0\|_{w,0} \\ &\leq \frac{1}{N\sqrt{R_0(0)}} \left\|\widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X f\right\|_{w,0} \\ &\leq \sqrt{R_0(0)} \|f\|_{w,0}.\end{aligned}$$

■

We now prove our approximation theorem:

Theorem 166 *Suppose $f, g \in X_w^0$ satisfy equation 4.54. Then*

$$\mathcal{S}_X^e f(x) - \frac{\left(\widetilde{\mathcal{E}}_X R_x\right)^T \widetilde{\mathcal{E}}_X f}{(R_0(0) + \rho) N} = \frac{R_0(0)}{R_0(0) + \rho} \mathcal{S}_X^e g(x), \quad x \in \mathbb{R}^d, \tag{4.58}$$

and

$$\left|\mathcal{S}_X^e f(x) - \frac{\left(\widetilde{\mathcal{E}}_X R_x\right)^T \widetilde{\mathcal{E}}_X f}{(R_0(0) + \rho) N}\right| \leq \frac{2R_0(0)^{3/2}}{R_0(0) + \rho} \min\left\{1, \frac{R_0(0)}{\rho}\right\} \|f\|_{w,0}, \tag{4.59}$$

and

$$\left\|\mathcal{S}_X^e f - \frac{\left(\widetilde{\mathcal{E}}_X R_x\right)^T \widetilde{\mathcal{E}}_X f}{(R_0(0) + \rho) N}\right\|_{w,0} \leq \frac{2R_0(0)}{R_0(0) + \rho} \|f\|_{w,0}. \tag{4.60}$$

Proof. Starting with 4.11 and using the fact that $\widetilde{\mathcal{E}}_X \widetilde{\mathcal{E}}_X^* = R_{X,X}$ (part 5 Theorem 52) we have

$$\begin{aligned}\mathcal{S}_X^e f(x) &= \left(\widetilde{\mathcal{E}}_X R_x\right)^T (\rho NI + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X f \\ &= \left(\widetilde{\mathcal{E}}_X R_x\right)^T (\rho NI + R_{X,X})^{-1} \left(\frac{1}{NR_0(0)} R_{X,X} \widetilde{\mathcal{E}}_X f + \widetilde{\mathcal{E}}_X g\right) \\ &= \frac{1}{NR_0(0)} \left(\widetilde{\mathcal{E}}_X R_x\right)^T (\rho NI + R_{X,X})^{-1} R_{X,X} \widetilde{\mathcal{E}}_X f + \left(\widetilde{\mathcal{E}}_X R_x\right)^T (\rho NI + R_{X,X})^{-1} \widetilde{\mathcal{E}}_X g \\ &= \frac{1}{NR_0(0)} \left(\widetilde{\mathcal{E}}_X R_x\right)^T (\rho NI + R_{X,X})^{-1} R_{X,X} \widetilde{\mathcal{E}}_X f + \mathcal{S}_X^e g(x) \\ &= \frac{1}{NR_0(0)} \left(\widetilde{\mathcal{E}}_X R_x\right)^T (\rho NI + R_{X,X})^{-1} (\rho NI + R_{X,X}) \widetilde{\mathcal{E}}_X f - \\ &\quad - \frac{1}{NR_0(0)} \left(\widetilde{\mathcal{E}}_X R_x\right)^T (\rho NI + R_{X,X})^{-1} \left(\rho N \widetilde{\mathcal{E}}_X f\right) + \mathcal{S}_X^e g(x) \\ &= \frac{\left(\widetilde{\mathcal{E}}_X R_x\right)^T \widetilde{\mathcal{E}}_X f}{NR_0(0)} - \frac{\rho}{R_0(0)} \mathcal{S}_X^e f(x) + \mathcal{S}_X^e g(x),\end{aligned} \tag{4.61}$$

and solving for $\mathcal{S}_X^e f(x)$ gives

$$\begin{aligned}\mathcal{S}_X^e f(x) &= \frac{1}{1+\frac{\rho}{R_0(0)}}\frac{\left(\widetilde{\mathcal{E}}_X R_x\right)^T \widetilde{\mathcal{E}}_X f}{NR_0(0)} + \frac{\mathcal{S}_X^e g(x)}{1+\frac{\rho}{R_0(0)}} \\ &= \frac{\left(\widetilde{\mathcal{E}}_X R_x\right)^T \widetilde{\mathcal{E}}_X f}{\left(R_0(0)+\rho\right)N} + \frac{R_0(0)}{R_0(0)+\rho}\mathcal{S}_X^e g(x),\end{aligned}$$

i.e.

$$\mathcal{S}_X^e f(x) - \frac{\left(\widetilde{\mathcal{E}}_X R_x\right)^T \widetilde{\mathcal{E}}_X f}{\left(R_0(0)+\rho\right)N} = \frac{R_0(0)}{R_0(0)+\rho}\mathcal{S}_X^e g(x),$$

which is equation 4.58. To prove the estimate 4.59 we will estimate $|\mathcal{S}_X^e g(x)|$ using the inequality 4.15:

$$|\mathcal{S}_X^e u(x)| \le \sqrt{R_0(0)}\min\left\{1, \frac{R_0(0)}{\rho}\right\}\|u\|_{w,0}, \quad u \in X_w^0. \tag{4.62}$$

Applying the Exact smoother to equation 4.55 gives

$$\mathcal{S}_X^e g(x) = \mathcal{S}_X^e f(x) - \frac{1}{NR_0(0)}\mathcal{S}_X^e \widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X f(x),$$

and then applying the estimate 4.62 to both terms on the right side yields

$$\begin{aligned}|\mathcal{S}_X^e g(x)| &\le |\mathcal{S}_X^e f(x)| + \frac{1}{NR_0(0)}\left|\mathcal{S}_X^e\left(\widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X f\right)(x)\right| \\ &\le \sqrt{R_0(0)}\min\left\{1, \frac{R_0(0)}{\rho}\right\}\|f\|_{w,0} + \frac{\sqrt{R_0(0)}}{N\|R_0\|_{w,0}^2}\min\left\{1, \frac{R_0(0)}{\rho}\right\}\left\|\widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X f\right\|_{w,0} \\ &= \sqrt{R_0(0)}\min\left\{1, \frac{R_0(0)}{\rho}\right\}\left(\|f\|_{w,0} + \frac{1}{NR_0(0)}\left\|\widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X f\right\|_{w,0}\right) \\ &\le 2\sqrt{R_0(0)}\min\left\{1, \frac{R_0(0)}{\rho}\right\}\|f\|_{w,0},\end{aligned}$$

where the last step used 4.57. Finally

$$\begin{aligned}\left|\mathcal{S}_X^e f(x) - \frac{\left(\widetilde{\mathcal{E}}_X R_x\right)^T \widetilde{\mathcal{E}}_X f}{\left(R_0(0)+\rho\right)N}\right| &= \frac{R_0(0)}{R_0(0)+\rho}|\mathcal{S}_X^e g(x)| \\ &\le \frac{2R_0(0)^{3/2}}{R_0(0)+\rho}\min\left\{1, \frac{R_0(0)}{\rho}\right\}\|f\|_{w,0},\end{aligned}$$

which proves 4.59. ■

Remark 167

1. *If we write*

$$\begin{aligned}\frac{\left(\widetilde{\mathcal{E}}_X R_x\right)^T \widetilde{\mathcal{E}}_X f}{\left(R_0(0)+\rho\right)N} &= \frac{1}{\left(R_0(0)+\rho\right)N}\sum_{k=1}^N f\left(x^{(k)}\right)R_x\left(x^{(k)}\right) \\ &= \frac{(2\pi)^{-d/2}}{\left(R_0(0)+\rho\right)}\frac{1}{N}\sum_{k=1}^N f\left(x^{(k)}\right)G\left(x-x^{(k)}\right) \\ &= \frac{1}{\left(G(0)+(2\pi)^{d/2}\rho\right)}\frac{1}{N}\sum_{k=1}^N f\left(x^{(k)}\right)G\left(x-x^{(k)}\right),\end{aligned}$$

it is clear that the mapping $f \to \frac{\left(\widetilde{\mathcal{E}}_X R_x\right)^T \widetilde{\mathcal{E}}_X f}{(R_0(0)+\rho)N}$ *is a continuous linear operator from* X_w^0 *to* $W_{G,X}$ *with the form of a weighted average, the weights being data-translated basis functions. Note also that the operator only depends on the values of the data function in the data region.*

2. *The approximation 4.59 provides a bound of order* ρ^{-2} *for large* ρ*. For all* ρ *this estimate supplies an upper bound of* $2\sqrt{R_0\left(0\right)}\left\|f\right\|_{w,0}$.

3. *The normwise approximation 4.60 provides a bound of order* ρ^{-1} *as* $\rho \to \infty$*. For all* ρ *this estimate supplies an upper bound of* $2\left\|f\right\|_{w,0}$.

5

The Approximate smoother

5.1 Introduction

We call this smoother the *Approximate smoother* because it approximates the Exact smoother. This is a non-parametric, scalable smoother. Here *scalable* means the numeric effort to calculate the Approximate smoother depends linearly on the number of data points. We assume the basis function is real-valued.

Two different approaches will be taken to defining the Approximate smoother, and both involve formulating the smoother as the solution of a variational problem. One of these problems will involve minimizing the Exact smoother functional 6 over $W_{G,X'}$ where $X' = \{x'_i\}_{i=1}^{N'}$ is an arbitrary set of distinct points in $\mathbb{R}^d$. The other, equivalent problem, involves finding the function in $W_{G,X'}$ which is nearest to Exact smoother s_e w.r.t. the norm $\|\cdot\|_{w,0}$. If

$$s_a(x) = \sum_{i=1}^{N'} \alpha'_i R_{x'_i}(x) = (2\pi)^{-d/2} \sum_{i=1}^{N'} \alpha'_i G(x - x'_i),$$

denotes the Approximate smoother and y is the dependent data then solving the second problem yields

$$s_a = \mathcal{I}_{X'} s_e,$$

which implies the matrix equation

$$\left(N\rho R_{X',X'} + R_{X,X'}^T R_{X,X'}\right)\alpha' = R_{X,X'}^T y,$$

where $R_{X,X'} = \left(R_{x'_j}\left(x^{(i)}\right)\right)$. The size of the Approximate smoother matrix is $N' \times N'$ which is independent of the number of data points and suggests scalability.

The error estimates for the pointwise convergence of the Approximate smoother to its data function $f \in X_w^0$ are based on the simple triangle inequality

$$|f(x) - s_a(x)| \le |f(x) - s_e(x)| + |s_e(x) - s_a(x)|,$$

and so Section 5.6 will be devoted to estimating $|s_e(x) - s_a(x)|$.

As with the minimal interpolant and the Exact smoother, we will obtain estimates that assume unisolvent data sets as well as the Type 1 and Type 2 estimates that do not involve unisolvency. The Approximate smoother convergence orders and the constants are the same as those for the interpolation case which are given in the interpolation tables 1, 2, 3 and 4.

Non-unisolvent data: Type 1 error estimates No *a priori* assumption is made concerning the weight function parameter κ but it will be assumed that the basis function satisfies an inequality of the form 1.

For example, if it is assumed that the data region K is closed, bounded and infinite then Theorem 189 establishes that

$$|s_e(x) - s_a(x)| \leq \|f\|_{w,0}\, k_G \left(h_{X',K}\right)^s, \quad x \in K,$$

and Theorem 206 shows that

$$|f(x) - s_a(x)| \leq \|f\|_{w,0} \left(\sqrt{\rho N} + k_G \left(h_{X,K}\right)^s + k_G \left(h_{X',K}\right)^s\right), \quad x \in K,$$

when $h_{X,K} = \max\limits_{x\in K} \operatorname{dist}(x,X) \leq h_G$ and $h_{X',K} = \max\limits_{x\in K} \operatorname{dist}(x,X') \leq h_G$.

Non-unisolvent data: Type 2 error estimates If it only assumed that $\kappa \geq 1$ then by Theorem 196

$$|s_e(x) - s_a(x)| \leq \|f\|_{w,0}\, k_G \left(h_{X',K}\right)^s, \quad x \in \mathbb{R}^d,$$

and by Theorem 208

$$|f(x) - s_a(x)| \leq \|f\|_{w,0} \left(\sqrt{\rho N} + k_G h_{X,K} + k_G h_{X',K}\right), \quad x \in \mathbb{R}^d,$$

where $k_G = (2\pi)^{-\frac{d}{4}} \sqrt{-\left(\nabla^2 G\right)(0)}\sqrt{d}$.

Unisolvent data error estimates If X is a unisolvent set of order $m \geq 1$ contained in an open, bounded data region Ω then by Theorem 200

$$|s_e(x) - s_a(x)| \leq \|f\|_{w,0}\, k_G \left(h_{X',\Omega}\right)^m, \quad x \in \overline{\Omega},$$

and by Theorem 211

$$|f(x) - s_a(x)| \leq \|f\|_{w,0} \left(K'_{\Omega,m}\sqrt{\rho N} + k_G \left(h_{X,K}\right)^m + k_G \left(h_{X',K}\right)^m\right), \quad x \in \overline{\Omega},$$

for some constants $K'_{\Omega,m}, k_G > 0$. We say the orders of convergence are at least m.

These theoretical error results will be illustrated using the weight function examples from the interpolation chapter, namely the radial *shifted thin-plate splines*, *Gaussian* and *Sobolev splines* and the tensor product *extended B-splines*. We will also use the *central difference* weight functions from Chapter 3.

Numerical results are only presented for the non-unisolvent data cases. Numeric experiments are carried out using the same 1-dimensional B-splines and data functions that were used for the interpolants. We restrict ourselves to one dimension so that the data density parameters $h_{X,\Omega}$ and $h_{X,K}$ can be easily calculated.

The last section discusses the *SmoothOperator* software (freeware) package which implements the Approximate smoother algorithm. It has a full user manual which describe several tutorials and data experiments.

5.2 The convolution space J_G

Following Dyn [3] we introduce the space J_G and prove some of its properties, including the fact that J_G is dense in X_w^0. This space will be used in the discretization process which derives the Approximate smoother problem from the Exact smoother problem in Subsection 5.3.1.

Definition 168 ***The space*** J_G

Suppose the weight function w satisfies property W2 for some $\kappa \geq 0$ and let G be the zero order basis function generated by w. Then

$$J_G = \{G * \phi : \phi \in S\},$$

where the convolution $G * \phi$ is defined by $G * \phi = \left(\widehat{\phi}\widehat{G}\right)^{\vee}$ with $G \in S'$ and $\phi \in S$. We will sometimes write $J_G = G * S$.

Theorem 169 *The function space J_G has the following properties:*

1. *$J_G \subset X_w^0 \cap C_B^\infty$.*
2. *If $f \in X_w^0$ then $(f, G * \phi)_{w,0} = \left[f, \overline{\phi}\right]$ for all $\phi \in S$.*
3. *J_G is dense in X_w^0.*

Proof. Suppose $g \in J_G$, say $g = G * \phi$ where $\phi \in S$.

Part 1. $\widehat{g} = \widehat{\phi}\,\widehat{G} = \frac{\widehat{\phi}}{w} \in L^1$ since property W2 implies $\frac{1}{w} \in L^1$. Thus

$$\int w\,|\widehat{g}|^2 = \int w \frac{\left|\widehat{\phi}\right|^2}{(w)^2} = \int \frac{\left|\widehat{\phi}\right|^2}{w} \leq \left\|\widehat{\phi}\right\|_\infty^2 \int \frac{1}{w} < \infty,$$

and so $g \in X_w^0$. For any multi-index $\alpha \geq 0$

$$(D^\alpha (G * \phi))^\wedge = i^{|\alpha|}\xi^\alpha \widehat{G * \phi} = i^{|\alpha|}\xi^\alpha \widehat{\phi}\widehat{G} = \frac{i^{|\alpha|}\xi^\alpha \widehat{\phi}}{w} \in L^1,$$

and so by Lemma 13, $D^\alpha (G * \phi) \in C_B^{(0)}$. Thus $G * \phi \in C_B^\infty$.

Part 2. $f \in X_w^0$ implies $\widehat{f} \in S' \cap L_{loc}^1$ and

$$(f, G * \phi)_{w,0} = \int w\widehat{f}\ \overline{\widehat{G * \phi}} = \int w\widehat{f}\frac{\overline{\widehat{\phi}}}{w} = \int \widehat{f}\overline{\widehat{\phi}}.$$

We now need some results from part (a) of Subsection 2.8.3 of Vladimirov [12]. Here he defines locally integrable functions which have polynomial growth at infinity i.e. a function $g \in L_{loc}^1$ such that $\int \frac{|g|}{(1+|\cdot|)^s}$ for some $s \geq 0$. Vladimirov states that these functions have the property that $g \in S'$ and $[g, \phi] = \int g\phi$ for $\phi \in S$.

Now, from the proof of Theorem 24, $\widehat{f} \in L^1$ and so $\widehat{f}$ has polynomial growth at infinity which implies

$$\int \widehat{f}\overline{\widehat{\phi}} = \left[\widehat{f}, \overline{\widehat{\phi}}\right] = \left[f, \overline{\phi}\right].$$

Part 3. A standard result is that a subspace of a Hilbert space is dense iff its orthogonal complement is $\{0\}$. In fact, if $\langle f, G * \phi\rangle_{w,0} = 0$ for all $\phi \in S$ then part 2 implies $f = 0$. ■

5.3 The Approximate smoother

From part 5 of Summary 174 the matrix equation for the Exact smoother is $(N\rho I + R_{X,X})\,\alpha = y$. The construction and solution of this system does not generate a scalable matrix algorithm since the size of the smoothing matrix is $N \times N$ i.e. its size depends on the number of data points whereas it will turn out that the matrix size for a scalable algorithm is independent of the number of data points. In this section we will overcome this limitation and derive the Approximate smoother problem by discretizing the Exact smoothing problem on a grid. This will be done by using the space $J_G = G * S$ which is dense in X_w^0 and was introduced in Definition 168. The space J_G is not necessary for the specification of the Approximate smoother problem but this is how I derived the Approximate smoother algorithm: I came across the convolution space J_G in Dyn's review paper [3] and decided to approximate its functions using a regular rectangular grid X' and the Trapezoidal rule. This led to functions in the finite dimensional basis function space $W_{G,X'}$. Minimizing the Exact smoother functional $J_e[\cdot]$ given by 4.1 over the functions in $W_{G,X'}$ yields the Approximate smoother matrix equation 5.13. The size of this matrix is independent of the number of data points and hence the construction and solution of the Approximate smoother matrix equation is a *scalable algorithm*. In fact, the size of the matrix is equal to the number of grid points and so increases exponentially with the number of dimensions. In practice we are limited to two or three dimensions.

In actual fact, the set X' will be generalized from a grid to any set of distinct points. This could be a sparse grid, for example.

5.3.1 **Approach 1** *an analog of the Exact smoothing problem*

In this subsection we will provide some justification for approximating the infinite dimensional Hilbert space X_w^0 by a finite dimensional subspace $W_{G,X'}$, where X' is a regular, rectangular grid of points in $\mathbb{R}^d$. The space $W_{G,X'}$ will be used to define the Approximate smoothing problem. The set X' will then be generalized to include any set of distinct points. This discretisation process turns out to be similar to that described in Garcke and Griebel [4].

Definition 170 ***A regular, rectangular grid in*** $\mathbb{R}^d$

Let the grid occupy a rectangle $R(a;b)$, which has left-most point $a \in \mathbb{R}^d$ and right-most point b. Suppose the grid has $\mathcal{N}' = (N_1', N_2', \ldots, N_d')$ points in each dimension and let $h \in \mathbb{R}^d$ denote the grid sizes.

Then $X' = \left\{x_\alpha' = a + h\alpha \mid \alpha \in \mathbb{Z}^d \text{ and } 0 \le \alpha < \mathcal{N}'\right\}$ is the set of grid points.

Let $N^{'}$ be the number of grid points so that $N^{'} = (\mathcal{N}')^{\mathbf{1}} = \mathcal{N}_1'\mathcal{N}_2' \ldots \mathcal{N}_d'$, and of course we have the constraint $\mathcal{N}^{'}h = b - a$.

By Theorem 169 the space $J_G = G * S$ is dense in X_w^0 and so we will approximate the functions in S using the rectangular grid X' defined above. Our analysis will be matrix-based so we choose an order for the grid points and set $X^{'} = \{x_n'\}_{n=1}^{N^{'}}$. Integrals on $\mathbb{R}^d$ will be approximated by integrals on the grid region using the trapezoidal rule i.e.

$$\int\limits_{grid} f(x)\, dx \simeq h^{\mathbf{1}} \sum_{n=1}^{N^{'}} f(x_n'), \tag{5.1}$$

where $h^{\mathbf{1}} = h_1 h_2 \times \ldots \times h_d$ is the volume of a grid element. This will be a two stage approximation: first a restriction to the grid rectangle and then an application of the trapezoidal rule.

Step1 Approximation of the functions in $J_G = G * S$ by functions in $W_{G,X'}$.

Suppose $\phi \in S$. Then the trapezoidal approximation 5.1 on the grid X' gives

$$G * \phi = \int\limits_{\mathbb{R}^d} G(\cdot - y)\, \phi(y)\, dy \simeq \sum_{n=1}^{N'} G(\cdot - x_n') \left(h^1 \phi(x_n')\right). \tag{5.2}$$

The equations and approximations 5.1 and 5.2 suggest we approximate the space $G * S$ by functions of the form $\sum\limits_{n=1}^{N'} G(x - x_n')\, \alpha_n$, $\alpha_n \in \mathbb{C}$ i.e. by functions in $W_{G,X'}$.

Step 2 With this motivation we could now specify a smoothing problem which we will call an Approximate smoothing problem. This would involve minimizing the Exact smoothing functional 4.1 over $W_{G,X'}$ where X' is a rectangular grid. However, since the space $W_{G,X'}$ is defined when X' is any set of distinct points, we will define the following more general problem:

The Approximate smoothing problem
Minimize the Exact smoothing functional $J_e[f]$ for $f \in W_{G,X'}$, where X' is a set of distinct points in $\mathbb{R}^d$.

(5.3)

5.3.2 **Approach 2**

In Theorem 166 it was shown that for any $f \in X_w^0$

$$\left| \mathcal{S}_X^e f(x) - \frac{\left(\widetilde{\mathcal{E}}_X R_x\right)^T \widetilde{\mathcal{E}}_X f}{(R_0(0) + \rho) N} \right| \le \frac{2R_0(0)^{3/2}}{R_0(0) + \rho} \min\left\{1, \frac{R_0(0)}{\rho}\right\} \|f\|_{w,0}, \quad x \in \mathbb{R}^d,$$

and it was noted in Remark 167 that $\frac{(\widetilde{\mathcal{E}}_X R_x)^T \widetilde{\mathcal{E}}_X f}{(R_0(0)+\rho)N} \in W_{G,X}$ and this approximation is bounded uniformly on $\mathbb{R}^d$. Of course $R_x(y) = (2\pi)^{d/2} G(x - y)$. Now suppose our independent data points X are always

contained in a bounded open data region Ω. Overlay Ω with a regular, rectangular grid and let $X' = \{x'_n\}_{n=1}^{N'} \subset \Omega$ denote the grid points in Ω. Further assume that G and f are Lipschitz continuous on $\mathbb{R}^d$.

Since f is Lipschitz continuous on $\mathbb{R}^d$ all the derivatives exist and are bounded a.e. Noting part 2 of Remark 78 we have $X_w^0 \subset H^{1,\infty}$, $G \in H^{1,\infty}$ and for each x, $R_x \in H^{1,\infty}$. Indeed

$$R_x f \in H^{1,\infty} \cap C_B^{(0)}. \tag{5.4}$$

Next we want to express $\left(\widetilde{\mathcal{E}}_X R_x\right)^T \widetilde{\mathcal{E}}_X f$ in terms of $\left(\widetilde{\mathcal{E}}_{X'} R_x\right)^T \widetilde{\mathcal{E}}_{X'} f$ and a remainder term and then estimate the remainder term. But

$$\frac{1}{N}\left(\widetilde{\mathcal{E}}_X R_x\right)^T \widetilde{\mathcal{E}}_X f = \frac{1}{N}\sum_{m=1}^{N} R_x\left(x^{(m)}\right) f\left(x^{(m)}\right) = \frac{1}{N}\sum_{k=1}^{N'} \sum_{x^{(m)} \in X'_k} R_x\left(x^{(m)}\right) f\left(x^{(m)}\right), \tag{5.5}$$

where

$$X'_k = \left\{x^{(m)} \in X : x^{(m)} - x'_k \in R\left[0; h/2\right)\right\},$$

means that $\{X'_k\}_{k=1}^{N'}$ partitions X.

The properties 5.4 of $R_x f$ mean that the distribution Taylor series expansion of Lemma 42 can now be brought into play to yield

$$R_x\left(x^{(m)}\right) f\left(x^{(m)}\right) = R_x\left(x'_k\right) f\left(x'_k\right) + \mathcal{R}_1\left(R_x f\right)\left(x'_k, x^{(m)} - x'_k\right),$$

and for $x^{(m)} \in X'_k$

$$\begin{aligned}
\left|\left(\mathcal{R}_1\left(R_x f\right)\right)\left(x'_k, x^{(m)} - x'_k\right)\right| &\le \sqrt{d}\max_{|\beta|=1}\left\|D^\beta\left(R_x f\right)\right\|_\infty \left|x^{(m)} - x'_k\right| \\
&\le \frac{\sqrt{d}}{2}\max_{|\beta|=1}\left\|D^\beta\left(R_x f\right)\right\|_\infty |h| \\
&\le \frac{\sqrt{d}}{2}\max_{|\beta|\le 1}\left\|D^\beta R_x\right\|_\infty \max_{|\beta|\le 1}\left\|D^\beta f\right\|_\infty |h|.
\end{aligned} \tag{5.6}$$

Thus

$$\begin{aligned}
\sum_{x^{(m)} \in X'_k} R_x\left(x^{(m)}\right) f\left(x^{(m)}\right) &= \sum_{x^{(m)} \in X'_k} R_x\left(x'_k\right) f\left(x'_k\right) + \sum_{x^{(m)} \in X'_k} \mathcal{R}_1\left(R_x f\right)\left(x'_k, x^{(m)} - x'_k\right) \\
&= N'_k R_x\left(x'_k\right) f\left(x'_k\right) + \sum_{x^{(m)} \in X'_k} \mathcal{R}_1\left(R_x f\right)\left(x'_k, x^{(m)} - x'_k\right),
\end{aligned}$$

and $\frac{1}{N}\left(\widetilde{\mathcal{E}}_X R_x\right)^T \widetilde{\mathcal{E}}_X f$ can now be written

$$\frac{1}{N}\left(\widetilde{\mathcal{E}}_X R_x\right)^T \widetilde{\mathcal{E}}_X f = \sum_{k=1}^{N'} R_x\left(x'_k\right)\frac{N'_k}{N} f\left(x'_k\right) + \frac{1}{N}\sum_{k=1}^{N'}\sum_{x^{(m)} \in X'_k} \mathcal{R}_1\left(R_x f\right)\left(x'_k, x^{(m)} - x'_k\right),$$

with

$$\sum_{k=1}^{N'} R_x\left(x'_k\right)\frac{N'_k}{N} f\left(x'_k\right) \in W_{G,X'}.$$

Consequently the remainder estimate 5.6 implies

$$\begin{aligned}
\left|\frac{1}{N}\left(\widetilde{\mathcal{E}}_X R_x\right)^T \widetilde{\mathcal{E}}_X f - \sum_{k=1}^{N'} R_x\left(x'_k\right)\frac{N'_k}{N} f\left(x'_k\right)\right| &\le \frac{1}{N}\sum_{k=1}^{N'}\sum_{x^{(m)} \in X'_k}\left|\mathcal{R}_1\left(R_x f\right)\left(x'_k, x^{(m)} - x'_k\right)\right| \\
&\le \frac{1}{N}\sum_{k=1}^{N'}\sum_{x^{(m)} \in X'_k}\frac{\sqrt{d}}{2}\max_{|\beta|\le 1}\left\|D^\beta R_x\right\|_\infty \max_{|\beta|\le 1}\left\|D^\beta f\right\|_\infty |h| \\
&= \frac{\sqrt{d}}{2}\max_{|\beta|\le 1}\left\|D^\beta R_x\right\|_\infty \max_{|\beta|\le 1}\left\|D^\beta f\right\|_\infty |h|,
\end{aligned} \tag{5.7}$$

giving a linear dependence on the grid size.

The last estimate can be combined with 4.58 to give

$$\left|\mathcal{S}_X^e f(x) - \frac{\sum\limits_{k=1}^{N'} R_x(x'_k) \frac{N'_k}{N} f(x'_k)}{R_0(0)+\rho}\right| \leq \frac{2R_0(0)^{3/2}}{R_0(0)+\rho} \min\left\{1, \frac{R_0(0)}{\rho}\right\} \|f\|_{w,0} +$$

$$+ \frac{\sqrt{d}}{2} \max_{|\beta|\leq 1} \left\|D^\beta R_x\right\|_\infty \max_{|\beta|\leq 1} \left\|D^\beta f\right\|_\infty |h|, \quad x \in \mathbb{R}^d, \tag{5.8}$$

so that as $N \to \infty$ and $|h| \to 0$ this approximation is bounded pointwise on $\mathbb{R}^d$.

Now if a weight function has $\kappa \geq 1$ then $X_w^0 \subset C_B^{(1)}$, $G \in C_B^{(2)}$ and so $D^\beta f$ and $D^\beta G$ are always Lipschitz continuous when $|\beta| \leq 1$. This certainly applies to the shifted thin-plate splines and the Gaussian.

We will now try to improve this approximation by replacing the vector $\left(\frac{N'_k}{N} f(x'_k)\right) (R_0(0)+\rho)^{-1}$ by an arbitrary complex vector $\alpha' \in \mathbb{C}^{N'}$ and then finding a uniform bound for $\left|\mathcal{S}_X^e f(x) - \left(\widetilde{\mathcal{E}}_{X'}^* \alpha'\right)(x)\right|$ when $x \in \mathbb{R}^d$ and $f \in X_w^0$. In fact we use the standard approach of writing

$$\left|\mathcal{S}_X^e f(x) - \left(\widetilde{\mathcal{E}}_{X'}^* \alpha'\right)(x)\right| = \left|\left(\mathcal{S}_X^e f - \widetilde{\mathcal{E}}_{X'}^* \alpha', R_x\right)_{w,0}\right| \leq \left\|\mathcal{S}_X^e f - \widetilde{\mathcal{E}}_{X'}^* \alpha'\right\|_{w,0} \|R_0\|,$$

and then show there is a unique α' which minimizes $\left\|\mathcal{S}_X^e f - \widetilde{\mathcal{E}}_{X'}^* \alpha'\right\|_{w,0}$. This is just the adjoint formulation of the minimum seminorm interpolation problem and the calculations are simplified since G is real.

Theorem 171 *If $g \in X_w^0$ then $R_{X,X}^{-1} \widetilde{\mathcal{E}}_{X'} g = \underset{\alpha' \in \mathbb{C}^{N'}}{\operatorname{argmin}} \left\|g - \widetilde{\mathcal{E}}_{X'}^* \alpha'\right\|_{w,0}$ is unique and $\widetilde{\mathcal{E}}_{X'}^* R_{X,X}^{-1} \widetilde{\mathcal{E}}_{X'} g = \mathcal{I}_{X'} g$.*

Proof. If $\alpha' \in \mathbb{C}^{N'}$ then

$$\begin{aligned}
\left\|g - \widetilde{\mathcal{E}}_{X'}^* \alpha'\right\|_{w,0}^2 &= \left(g - \widetilde{\mathcal{E}}_{X'}^* \alpha', g - \widetilde{\mathcal{E}}_{X'}^* \alpha'\right)_{w,0} \\
&= \|g\|_{w,0}^2 - \left(g, \widetilde{\mathcal{E}}_{X'}^* \alpha'\right)_{w,0} - \left(\widetilde{\mathcal{E}}_{X'}^* \alpha', g\right)_{w,0} + \left(\widetilde{\mathcal{E}}_{X'}^* \alpha', \widetilde{\mathcal{E}}_{X'}^* \alpha'\right)_{w,0} \\
&= \|g\|_{w,0}^2 - 2\operatorname{Re}\left(g, \widetilde{\mathcal{E}}_{X'}^* \alpha'\right)_{w,0} + \left(\widetilde{\mathcal{E}}_{X'}^* \alpha', \widetilde{\mathcal{E}}_{X'}^* \alpha'\right)_{w,0} \\
&= \|g\|_{w,0}^2 - 2\operatorname{Re}\left(\widetilde{\mathcal{E}}_{X'} g, \alpha'\right) + \left(\alpha', \widetilde{\mathcal{E}}_{X'} \widetilde{\mathcal{E}}_{X'}^* \alpha'\right) \\
&= \|g\|_{w,0}^2 - 2\operatorname{Re}\left(\widetilde{\mathcal{E}}_{X'} g, \alpha'\right) + \left(\alpha', R_{X,X} \alpha'\right),
\end{aligned}$$

and if we substitute $\alpha' = \beta' + i\gamma'$ and differentiate w.r.t. β' and then γ' we obtain a unique minimum at $\alpha' = R_{X,X}^{-1} \widetilde{\mathcal{E}}_{X'} g$. ∎

This theorem implies that the closest point in $W_{G,X'}$ to $\mathcal{S}_X^e f$ under $\|\cdot\|_{w,0}$ is $\mathcal{I}_{X'} \mathcal{S}_X^e f$, and so we define the Approximate smoother operator $\mathcal{S}_X^a$ by

$$\mathcal{S}_X^a = \mathcal{I}_{X'} \mathcal{S}_X^e \tag{5.9}$$

It was observed that the Approximate smoother problem 5.3 still makes sense if X' is any set of distinct points in $\mathbb{R}^d$ and clearly Definition 5.9 is still meaningful if X' is any set of distinct points. In part 2 of Corollary 176 we will show that the solution of the Approximate smoother problem is the Approximate smoother 5.9.

5.4 Preparation: the Exact smoother mapping and data functions

This section presents a convenient summary of the Exact smoother mapping properties required for our study of the Approximate smoother problem 5.3.

The Exact smoother problem was described in the Introduction and was solved in Chapter 4 using a geometric Hilbert space framework which involved the introduction of the Hilbert product space V and the operator $\mathcal{L}_X : X_w^0 \to V$:

Definition 172 ***The space*** V ***and the operator*** $\mathcal{L}_X$

1. *Let* $V = X_w^0 \otimes \mathbb{C}^N$ *be the Hilbert product space with norm* $\|\cdot\|_V$ *and inner product* $(\cdot,\cdot)_V$ *given by*

$$((u_1,u_2),(v_1,v_2))_V = \rho\,(u_1,v_1)_{w,0} + \frac{1}{N}(u_2,v_2)_{\mathbb{C}^N}.$$

2. *Let the operator* $\mathcal{L}_X : X_w^0 \to V$ *be defined by* $\mathcal{L}_X f = \left(f, \widetilde{\mathcal{E}}_X f\right)$ *where* $\widetilde{\mathcal{E}}_X$ *is the vector-valued evaluation function* $\widetilde{\mathcal{E}}_X f = \left(f\left(x^{(k)}\right)\right)$ *of Definition 51.*

The Exact smoothing functional J_e can now be written

$$J_e[f] = \|\mathcal{L}_X f - (0,y)\|_V^2, \quad f \in X_w^0, \tag{5.10}$$

and the Exact smoother s_e is the unique orthogonal projection of $(0,y)$ onto the infinite dimensional subspace $\mathcal{L}_X\left(X_w^0\right)$. Using this approach it was shown in part 4 of Theorem 135 that

$$s_e = \frac{1}{N}\left(\mathcal{L}_X^*\mathcal{L}_X\right)^{-1}\widetilde{\mathcal{E}}_X^* y, \quad y \in \mathbb{C}^N. \tag{5.11}$$

Definition 173 ***Data functions and the Exact smoother mapping*** $\mathcal{S}_X^e : X_w^0 \to W_{G,X}$.
Given an independent data set X *each member of* X_w^0 *is assumed to act as a legitimate data function* f *and generate a dependent data vector* $\widetilde{\mathcal{E}}_X f$*. Further, equation 5.11 enables us to define a mapping* $\mathcal{S}_X^e : X_w^0 \to W_{G,X}$ *from the data functions to their corresponding Exact smoothers defined by*

$$\mathcal{S}_X^e f = \frac{1}{N}\left(\mathcal{L}_X^*\mathcal{L}_X\right)^{-1}\widetilde{\mathcal{E}}_X^*\widetilde{\mathcal{E}}_X f, \quad f \in X_w^0. \tag{5.12}$$

By Corollary 137, $\mathcal{L}_X^*\mathcal{L}_X$ *is a homeomorphism from* X_w^0 *to* X_w^0 *so* $\mathcal{S}_X^e$ *is a continuous linear mapping.*

Summary 174 *This is a list of some of the properties of the Exact smoother mapping* $\mathcal{S}_X^e$ *which will be used in this document:*

1. $\mathcal{S}_X^e$ *maps* X_w^0 *onto* $W_{G,X}$ *and* $\operatorname{null}\mathcal{S}_X^e = W_{G,X}^{\perp}$*. Also,* $\mathcal{S}_X^e$ *is self-adjoint but not a projection.*

2. $\|\mathcal{S}_X^e f\|_{w,0} \le \|f\|_{w,0}$ *and* $\|(I-\mathcal{S}_X^e) f\|_{w,0} \le \|f\|_{w,0}$, $\quad f \in X_w^0$.

3. $\mathcal{S}_X^e f = f - \rho\left(\mathcal{L}_X^*\mathcal{L}_X\right)^{-1} f, \quad f \in X_w^0$.

4. $\mathcal{S}_X^e f = \widetilde{\mathcal{E}}_X^*\left(N\rho I + R_{X,X}\right)^{-1}\widetilde{\mathcal{E}}_X f, \quad f \in X_w^0$.

5. $\mathcal{S}_X^e f = \sum\limits_{k=1}^{N} \alpha_k R_{x^{(k)}}$, *where* $\alpha = (\alpha_k)$ *satisfies the matrix equation*

   $\left(N\rho I + R_{X,X}\right)\alpha = y$, *where* $y = \widetilde{\mathcal{E}}_X f$ *is the dependent data.*

Proof. We just list the references from Chapter 4:

Part 1 Corollary 137; **Part 2** Part 2 of Theorem 140; **Part 3** Corollary 137; **Parts 4** and **5** Use part 3 of Theorem 138. ■

5.5 Properties of the Approximate smoother problem

It will be shown in part 1 of the next theorem that the Approximate smoothing problem has a unique solution, which we will call the Approximate smoother.

Theorem 175 *Denote the data for the Approximate smoothing problem by* $[X; y]$ *and* X'. *Set* $\varsigma = (0, y)$ *so that by 5.10,* $J_e[f] = \|\mathcal{L}_X f - \varsigma\|_V^2$ *is the Exact smoothing functional.*

Then the Approximate smoothing problem has a unique solution $s_a \in W_{G,X'}$ *which satisfies:*

1. $J_e[s_a] < J_e[f]$, *for all* $f \in W_{G,X'}$ *and* $f \neq s_a$.
2. $(\mathcal{L}_X s_a - \varsigma, \mathcal{L}_X s_a - \mathcal{L}_X f)_V = 0$, *for all* $f \in W_{G,X'}$.
3. $\|\mathcal{L}_X s_a - \varsigma\|_V^2 + \|\mathcal{L}_X s_a - \mathcal{L}_X f\|_V^2 = \|\mathcal{L}_X f - \varsigma\|_V^2$, *for all* $f \in W_{G,X'}$.

   Note that the equations of parts 2 and 3 are directly equivalent.

4. $\left(\mathcal{L}_X^* \mathcal{L}_X s_a - \frac{1}{N}\widetilde{\mathcal{E}}_X^* y,\ f\right)_{w,0} = 0$, *for all* $f \in W_{G,X'}$.

Proof. Part 1. We want to show that there is a unique function in $W_{G,X'}$ which minimizes the Exact smoothing functional $J_e[\cdot]$ over $W_{G,X'}$. Since $W_{G,X'}$ is a finite dimensional subspace of X_w^0, its image under $\mathcal{L}_X$ must be a finite dimensional subspace of V and hence a closed subspace of V.

Consequently, there exists a unique element of $\mathcal{L}_X(W_{G,X'})$, say v, which is the projection of ζ onto $\mathcal{L}_X(W_{G,X'})$, such that $\|v - \varsigma\|_V < \|\mathcal{L}_X f - \varsigma\|_V$ for all $f \in W_{G,X'}$ such that $\mathcal{L}_X(f) \neq v$.

Since $\mathcal{L}_X$ is 1-1 on X_w^0 there exists a unique element of $W_{G,X'}$, call it s_a, such that $v = \mathcal{L}_X(s_a)$.

In terms of J_e we have $J_e[s_a] < J_e[f]$ for all $f \in W_{G,X'}$ and $f \neq s_a$.

Parts 2 and 3. Since v is the projection of ζ onto $\mathcal{L}_X(W_{G,X'})$ simple Hilbert space geometry yields the equivalent equations of parts 2 and 3.

Part 4. We start with the equation of part 2 and use the fact from part 3 of Theorem 134 that $\mathcal{L}_X^* u = \rho u_1 + \frac{1}{N}\widetilde{\mathcal{E}}_X^* u_2$, $(u_1, u_2) \in V$. Noting that X has N data points we have for all $f \in W_{G,X'}$

$$\begin{aligned} 0 = (\mathcal{L}_X s_a - \varsigma, \mathcal{L}_X s_a - \mathcal{L}_X f)_V &= (\mathcal{L}_X s_a - \varsigma, \mathcal{L}_X(s_a - f))_V \\ &= (\mathcal{L}_X^* \mathcal{L}_X s_a - \mathcal{L}_X^* \varsigma, s_a - f)_{w,0} \\ &= \left(\mathcal{L}_X^* \mathcal{L}_X s_a - \frac{1}{N}\widetilde{\mathcal{E}}_X^* y, s_a - f\right)_{w,0}. \end{aligned}$$

But $s_a \in W_{G,X'}$ so

$$\left(\mathcal{L}_X^* \mathcal{L}_X s_a - \frac{1}{N}\widetilde{\mathcal{E}}_X^* y, f\right)_{w,0} = 0, \quad f \in W_{G,X'},$$

and thus $\mathcal{L}_X^* \mathcal{L}_X s_a - \frac{1}{N}\widetilde{\mathcal{E}}_X^* y \in W_{G,X'}^\perp$ by definition of the orthogonal complement $W_{G,X'}^\perp$. ■

Part 4 is used to prove the next two corollaries.

Corollary 176 *Suppose* s_e *is the Exact smoother of the data* $[X, y]$. *Suppose* s_a *is the Approximate smoother of the data* $[X, y]$ *generated by the points* X'. *Then:*

1. $s_e - s_a \in W_{G,X'}^\perp$ *i.e.* $\widetilde{\mathcal{E}}_{X'} s_e = \widetilde{\mathcal{E}}_{X'} s_a$.
2. $s_a = \widetilde{\mathcal{E}}_{X'}^* \left(\widetilde{\mathcal{E}}_{X'} \widetilde{\mathcal{E}}_{X'}^*\right)^{-1} \widetilde{\mathcal{E}}_{X'} s_e = \widetilde{\mathcal{E}}_{X'}^* R_{X',X'}^{-1} \widetilde{\mathcal{E}}_{X'} s_e = \mathcal{I}_{X'} s_e$.
3. $s_a = \frac{1}{N}\widetilde{\mathcal{E}}_{X'}^* R_{X',X'}^{-1} \widetilde{\mathcal{E}}_{X'} \left(\mathcal{L}_X^* \mathcal{L}_X\right)^{-1} \widetilde{\mathcal{E}}_X^* y$.

Proof. Part 1. From the proof of part 4 of the previous theorem $\mathcal{L}_X^* \mathcal{L}_X s_a - \frac{1}{N}\widetilde{\mathcal{E}}_X^* y \in W_{G,X'}^\perp$ so $\mathcal{L}_X^* \mathcal{L}_X s_a - \frac{1}{N}\widetilde{\mathcal{E}}_X^* y = g$ for some $g \in W_{G,X'}^\perp$ and $s_a - \frac{1}{N}\left(\mathcal{L}_X^* \mathcal{L}_X\right)^{-1} \widetilde{\mathcal{E}}_X^* y = \left(\mathcal{L}_X^* \mathcal{L}_X\right)^{-1} g$. But by equation 5.12, $s_e = \frac{1}{N}\left(\mathcal{L}_X^* \mathcal{L}_X\right)^{-1} \widetilde{\mathcal{E}}_X^* y$ and so $s_a - s_e = \left(\mathcal{L}_X^* \mathcal{L}_X\right)^{-1} g$.

Further, from part 3 of Summary 174, $\mathcal{S}_X^e g = g - \rho \left(\mathcal{L}_X^* \mathcal{L}_X\right)^{-1} g$ and thus $\left(\mathcal{L}_X^* \mathcal{L}_X\right)^{-1} g = \frac{1}{\rho} g$ since $\operatorname{null} \mathcal{S}_X^e = W_{G,X'}^{\perp}$. Hence $s_a - s_e = \frac{1}{\rho} g \in W_{G,X'}^{\perp}$ which is characterized by $\widetilde{\mathcal{E}}_{X'} (s_e - s_a) = 0$ i.e. $\widetilde{\mathcal{E}}_{X'} s_e = \widetilde{\mathcal{E}}_{X'} s_a$.

Part 2. From Theorem 52 we know $\widetilde{\mathcal{E}}_X^* : \mathbb{C}^N \to W_{G,X}$ and $\widetilde{\mathcal{E}}_{X'}^* : \mathbb{C}^{N'} \to W_{G,X'}$ are onto. Hence, since $s_a \in W_{G,X'}$ and $s_e \in W_{G,X}$ we have $s_e = \widetilde{\mathcal{E}}_X^* \alpha$ and $s_a = \widetilde{\mathcal{E}}_{X'}^* \beta$ for some $\alpha \in \mathbb{C}^N$ and $\beta \in \mathbb{C}^{N'}$. Parts 1 and 5 of Theorem 52 now imply

$$0 = \widetilde{\mathcal{E}}_{X'} (s_e - s_a) = \widetilde{\mathcal{E}}_{X'} \left(s_e - \widetilde{\mathcal{E}}_{X'}^* \beta\right) = \widetilde{\mathcal{E}}_{X'} s_e - \widetilde{\mathcal{E}}_{X'} \widetilde{\mathcal{E}}_{X'}^* \beta = \widetilde{\mathcal{E}}_{X'} s_e - R_{X',X'} \beta,$$

so that $\beta = R_{X',X'}^{-1} \widetilde{\mathcal{E}}_{X'} s_e$ and $s_a = \widetilde{\mathcal{E}}_{X'}^* \beta = \widetilde{\mathcal{E}}_{X'}^* R_{X',X'}^{-1} \widetilde{\mathcal{E}}_{X'} s_e$, as required.

The last equality follows from the formula for the minimal norm interpolant operator $\mathcal{I}_{X'}$ given by equation 2.11, namely $\mathcal{I}_{X'} f = \widetilde{\mathcal{E}}_{X'}^* \left(R_{X',X'}\right)^{-1} \widetilde{\mathcal{E}}_{X'} f$.

Part 3. From the proof of part 1, $s_e = \frac{1}{N} \left(\mathcal{L}_X^* \mathcal{L}_X\right)^{-1} \widetilde{\mathcal{E}}_X^* y$. Hence by part 2

$$\begin{aligned} s_a = \widetilde{\mathcal{E}}_{X'}^* R_{X',X'}^{-1} \widetilde{\mathcal{E}}_{X'} s_e &= \widetilde{\mathcal{E}}_{X'}^* R_{X',X'}^{-1} \widetilde{\mathcal{E}}_{X'} \frac{1}{N} \left(\mathcal{L}_X^* \mathcal{L}_X\right)^{-1} \widetilde{\mathcal{E}}_X^* y \\ &= \frac{1}{N} \widetilde{\mathcal{E}}_{X'}^* R_{X',X'}^{-1} \widetilde{\mathcal{E}}_{X'} \left(\mathcal{L}_X^* \mathcal{L}_X\right)^{-1} \widetilde{\mathcal{E}}_X^* y. \end{aligned}$$

■

The following corollary contains results which are analogous to some of those derived in Theorem 140 for the Exact smoother.

Corollary 177 *Suppose $s_a \in W_{G,X'}$ is the (unique) Approximate smoother of the data $[X, y]$ generated by the points X'. Then:*

1. $(s_a, f)_{w,0} = \frac{1}{N\rho} \left(y - (s_a)_X\right)^T \overline{f_X}$, *when* $f \in W_{G,X'}$.
2. $\|s_a\|_{w,0}^2 = \frac{1}{N\rho} \left(y - (s_a)_X\right)^T \overline{(s_a)_X}$.
3. $J_e(s_a) = \frac{1}{N} \left(y - (s_a)_X\right)^T \overline{y}$.

Proof. Part 1. From the definition of $\mathcal{L}_X^* \mathcal{L}_X$

$$\mathcal{L}_X^* \mathcal{L}_X s_a - \frac{1}{N} \widetilde{\mathcal{E}}_X^* y = \rho s_a + \frac{1}{N} \widetilde{\mathcal{E}}_X^* \widetilde{\mathcal{E}}_X s_a - \frac{1}{N} \widetilde{\mathcal{E}}_X^* y = \rho s_a + \frac{1}{N} \widetilde{\mathcal{E}}_X^* \left((s_a)_X - y\right).$$

Thus, when $f \in W_{G,X'}$

$$\begin{aligned} 0 = \left(\mathcal{L}_X^* \mathcal{L}_X s_a - \frac{1}{N} \widetilde{\mathcal{E}}_X^* y, f\right)_{w,0} &= (\rho s_a, f)_{w,0} + \frac{1}{N} \left(\widetilde{\mathcal{E}}_X^* \left((s_a)_X - y\right), f\right)_{w,0} \\ &= \rho (s_a, f)_{w,0} + \frac{1}{N} \left((s_a)_X - y, \widetilde{\mathcal{E}}_X f\right)_{\mathbb{C}^N} \\ &= \rho (s_a, f)_{w,0} + \frac{1}{N} \left((s_a)_X - y, f_X\right)_{\mathbb{C}^N} \\ &= \rho (s_a, f)_{w,0} + \frac{1}{N} \left((s_a)_X - y\right)^T \overline{f_X}. \end{aligned}$$

Part 2. If $f = s_a$ in the equation proved in part 1, then

$$0 = \rho \|s_a\|_{w,0}^2 + \frac{1}{N} \left((s_a)_X - y, (s_a)_X\right)_{\mathbb{C}^N} = \rho \|s_a\|_{w,0}^2 + \frac{1}{N} \left((s_a)_X - y\right)^T \overline{(s_a)_X},$$

so that $\|s_a\|_{w,0}^2 = \frac{1}{\rho N} \left(y - (s_a)_X\right)^T \overline{(s_a)_X}$.

Part 3.

$$
\begin{aligned}
J_e(s) &= \rho\left\|s_a\right\|_{w,0}^2 + \frac{1}{N}\left|(s_a)_X - y\right|^2 \\
&= \rho\left\|s_a\right\|_{w,0}^2 + \frac{1}{N}\left((s_a)_X^T - y^T\right)\left(\overline{(s_a)_X} - \overline{y}\right) \\
&= \rho\left\|s_a\right\|_{w,0}^2 + \frac{1}{N}\left((s_a)_X^T\,\overline{(s_a)_X} - (s_a)_X^T\,\overline{y} - y^T\overline{(s_a)_X} + y^T\overline{y}\right) \\
&= \frac{1}{N}\left(y - (s_a)_X\right)^T\overline{(s_a)_X} + \frac{1}{N}\left((s_a)_X^T\,\overline{(s_a)_X} - (s_a)_X^T\,\overline{y} - y^T\overline{(s_a)_X} + y^T\overline{y}\right) \\
&= \frac{1}{N}\left(y^T\overline{(s_a)_X} - (s_a)_X^T\,\overline{(s_a)_X} + (s_a)_X^T\,\overline{(s_a)_X} - (s_a)_X^T\,\overline{y} - y^T\overline{(s_a)_X} + y^T\overline{y}\right) \\
&= \frac{1}{N}\left(-(s_a)_X^T\,\overline{y} + y^T\overline{y}\right) \\
&= \frac{1}{N}\left(y - (s_a)_X\right)^T\overline{y}.
\end{aligned}
$$

■

In a manner analogous to the minimal norm interpolant mapping $\mathcal{I}_X$ of Definition 55 and the Exact smoother mapping $\mathcal{S}_X^e$ of Definition 136 which map data functions to their interpolant and smoother respectively, we will now define the Approximate smoother mapping:

Definition 178 ***The Approximate smoother mapping*** $\mathcal{S}_{X,X'}^a : X_w^0 \to W_{G,X'}$

Given an independent data set X, we shall assume that each member of X_w^0 can act as a legitimate data function f_d and generate the data vector $\widetilde{\mathcal{E}}_X f_d$.

The equation of part 2 of Corollary 176 enables us to define a continuous linear mapping $\mathcal{S}_{X,X'}^a$: $X_w^0 \to W_{G,X'}$ from the data functions to the corresponding Approximate smoother of the data $\left[X, \widetilde{\mathcal{E}}_X f_d\right]$ generated by the points X'. This mapping is given by

$$
\mathcal{S}_{X,X'}^a f_d = \frac{1}{N}\widetilde{\mathcal{E}}_{X'}^* R_{X',X'}^{-1}\widetilde{\mathcal{E}}_{X'}\left(\mathcal{L}_X^*\mathcal{L}_X\right)^{-1}\widetilde{\mathcal{E}}_X^*\widetilde{\mathcal{E}}_X f_d, \quad f_d \in X_w^0.
$$

We now prove some properties of the Approximate smoother mapping.

Corollary 179 *Suppose $\mathcal{S}_{X,X'}^a$ is the Approximate smoother mapping. Then $\mathcal{S}_{X,X'}^a = \mathcal{I}_{X'}\mathcal{S}_X^e$ where $\mathcal{I}_{X'}$ is the minimal norm interpolant mapping with independent data X' and $\mathcal{S}_X^e$ is the Exact smoother mapping with independent data X.*

Further, $\mathcal{S}_{X,X'}^a$ is self-adjoint and $\left\|\mathcal{S}_{X,X'}^a f\right\|_{w,0} \le \|f\|_{w,0}$ for all $f \in X_w^0$.

Proof. That $\mathcal{S}_{X,X'}^a = \mathcal{I}_{X'}\mathcal{S}_X^e$ follows immediately from 5.12 for $\mathcal{S}_X^e$ and equation 2.11 for $\mathcal{I}_{X'}$. $\mathcal{S}_{X,X'}^a$ is self-adjoint since $\mathcal{I}_{X'}$ and $\mathcal{S}_X^e$ are both self-adjoint. Finally, from part 2 Summary 174, $\left\|\mathcal{S}_{X,X'}^a f\right\|_{w,0} \le \|f\|_{w,0}$ since $\|\mathcal{S}_X^e f\|_{w,0} \le \|f\|_{w,0}$ and $\|\mathcal{I}_{X'} f\|_{w,0} \le \|f\|_{w,0}$ for all $f \in X_w^0$. ■

In the last corollary it was shown that the Approximate smoother is an interpolant of the Exact smoother. Hence the Approximate smoother uses the same information about the data function as the Exact smoother does i.e. it involves no additional data function evaluations.

5.5.1 The matrix equation for the Approximate smoother

We now know the Approximate smoother exists and is unique. The next step is to derive its matrix equation.

Theorem 180 *Suppose $s \in W_{G,X'}$ is the (unique) Approximate smoother of the data $[X, y]$ generated by the distinct points $X' = \{x_i'\}_{i=1}^{N'}$.*

Then $s(x) = \sum_{i=1}^{N'} \alpha_i' R_{x_i'}(x)$, where $\alpha' = (\alpha_i') \in \mathbb{C}^{N'}$ and α' satisfies

$$
\left(N\rho R_{X',X'} + R_{X,X'}^T R_{X,X'}\right)\alpha' = R_{X,X'}^T y, \tag{5.13}
$$

where $R_{X,X} = (2\pi)^{-\frac{d}{2}} G_{X,X}$ *and* $R_{X,X'} = (2\pi)^{-\frac{d}{2}} G_{X,X'}$.

The matrix on the left of equation 5.13 will be called the Approximate smoother matrix and it will usually be denoted by the symbol Ψ.

Proof. This proof uses the results $\widetilde{\mathcal{E}}_{X'}\widetilde{\mathcal{E}}_X^* = R_{X',X}$ and $\widetilde{\mathcal{E}}_X\widetilde{\mathcal{E}}_{X'}^* = R_{X,X'}$ which follow from part 6 of Theorem 52. From Theorem 175, $\left(\mathcal{L}_X^*\mathcal{L}_X s - \frac{1}{N}\widetilde{\mathcal{E}}_X^* y,\, f\right)_{w,0} = 0$ for all $f \in W_{G,X'}$. Substituting $f = R_{x_i'}$ yields $\widetilde{\mathcal{E}}_{X'}\left(\mathcal{L}_X^*\mathcal{L}_X s - \frac{1}{N}\widetilde{\mathcal{E}}_X^* y\right) = 0$, and since by part 4 Theorem 134, $\mathcal{L}_X^*\mathcal{L}_X s = \rho s + \frac{1}{N}\widetilde{\mathcal{E}}_X^*\widetilde{\mathcal{E}}_X s$, we have

$$\begin{aligned}0 = \widetilde{\mathcal{E}}_{X'}\left(\mathcal{L}_X^*\mathcal{L}_X s - \frac{1}{N}\widetilde{\mathcal{E}}_X^* y\right) &= \widetilde{\mathcal{E}}_{X'}\left(\rho s + \frac{1}{N}\widetilde{\mathcal{E}}_X^*\widetilde{\mathcal{E}}_X s - \frac{1}{N}\widetilde{\mathcal{E}}_X^* y\right)\\ &= \rho\widetilde{\mathcal{E}}_{X'} s + \frac{1}{N}R_{X',X}\widetilde{\mathcal{E}}_X s - \frac{1}{N}R_{X',X}y\\ &= N\rho s_{X'} + R_{X',X}s_X - R_{X',X}y. \end{aligned} \tag{5.14}$$

The next step is to write $s_{X'}$ and s_X in terms of basis functions. But $s(x) = \sum_{i=1}^{N'} \alpha_i' R_{x_i'}(x) = \widetilde{\mathcal{E}}_{X'}^*\alpha'$ implies

$$s_X = R_{X,X'}\alpha', \qquad s_{X'} = R_{X',X'}\alpha', \tag{5.15}$$

so that 5.14 becomes

$$0 = N\rho R_{X',X'}\alpha' + R_{X',X}R_{X,X'}\alpha' - R_{X',X}y,$$

which can be rearranged to give the desired matrix equation 5.13. ■

Remark 181

1. *Note that the reason the matrix equation 5.13 is more concise written in matrices based on* R_x *than the basis function* G *is that I have chosen the 'symmetric' Fourier transform pair* $\widehat{f}(\xi) = (2\pi)^{-\frac{d}{2}}\int e^{-ix\xi}f(x)dx$ *and* $\overset{\vee}{f}(\xi) = (2\pi)^{-\frac{d}{2}}\int e^{ix\xi}f(x)dx$.

2. *The 'basis function form' of the Approximate smoother matrix equation 5.13 is*

$$\left((2\pi)^{\frac{d}{2}} N\rho G_{X',X'} + G_{X,X'}^T G_{X,X'}\right)\alpha' = G_{X,X'}^T y,$$

with the smoother given by

$$s(x) = \sum_{i=1}^{N'} \alpha_i' G(x - x_i').$$

In the next theorem we prove some properties of the Approximate smoother matrix.

Theorem 182 *The Approximate smoother matrix* Ψ *specified in Theorem 180 has the following properties:*

1. Ψ *is real valued, symmetric, positive definite and regular.*

2. Ψ *has size* $N' \times N'$. *Hence the size of* Ψ *is independent of the number of (scattered) data points.*

Proof. Part 1. Since G is a Hermitian function and is assumed to be real valued in this document, it follows that $R_{X',X'}$ and $R_{X',X}$ have real elements, $R_{X',X'}$ is symmetric and so Ψ is real and symmetric. Further, if $\alpha \in \mathbb{C}^N$

$$\begin{aligned}\alpha^T\Psi\overline{\alpha} = \alpha^T\left(N\rho R_{X',X'} + R_{X,X'}^T R_{X,X'}\right)\overline{\alpha} &= N\rho\alpha^T R_{X',X'}\overline{\alpha} + \alpha^T R_{X,X'}^T R_{X,X'}\overline{\alpha}\\ &= N\rho\alpha^T R_{X',X'}\overline{\alpha} + \left|R_{X,X'}\overline{\alpha}\right|^2.\end{aligned}$$

But from the Introduction 5.1 $R_{X',X'}$ is positive definite and so $\alpha^T\Psi\overline{\alpha} > 0$ iff $\alpha = 0$ which implies that Ψ is positive definite over $\mathbb{C}$, and since G is real we have that Ψ is regular.

Part 2. From the block sizes it is clear that Ψ is square with $N' + 2M$ rows. Hence the size of Ψ is independent of the number of (scattered) data points N. ■

5.5.2 The scalability of the Approximate smoother

From the basis function G, the N independent data points X and the N' points X', we construct the $N' \times N'$ matrix $R_{X',X'} = (2\pi)^{-\frac{d}{2}} G_{X',X'}$ and the $N \times N'$ matrix $R_{X,X'} = (2\pi)^{-\frac{d}{2}} G_{X,X'}$. For a specified smoothing parameter ρ and independent data y, we construct the matrix equation 5.13 i.e.

$$\left(N\rho R_{X',X'} + R_{X,X'}^T R_{X,X'} \right) \alpha' = R_{X,X'}^T y.$$

The system is solved for α' and the Approximate smoother is evaluated at various points Z using the formula $s_Z = R_{Z,X'}\alpha'$.

Our next result shows the algorithm is *scalable* i.e. the time of construction and execution of the solution is proportional to the number of data points. This is in contrast with the Exact smoother which is not scalable but which has quadratic dependency on the number of data points.

Corollary 183 *The Approximate smoother algorithm is scalable.*

Proof. Suppose the *evaluation cost* for $G(x)$ is m_G multiplications and that $N \gg N'$. The *construction cost* for Ψ is

$$(N')^2 m_G + N'Nm_G + N'N + N'Nm_G \simeq 2N'Nm_G.$$

The *solution cost* of an $N' \times N'$ matrix equation is $\frac{1}{3}(N')^3$ multiplications for a dense matrix. Thus the *total cost* is $2N'Nm_G + \frac{1}{3}(N')^3$ which is linearly dependent on the number of data points.

However, by the use of a basis function with support containing only several points in X' e.g. the hat function, the construction and solution costs can be reduced significantly. However we still have linear dependency on N. ■

5.6 Convergence of the Approximate smoother to the Exact smoother

In this Section we will prove several results concerning the convergence of the Exact smoother to the Approximate smoother. We will start by proving some general results which only assume that the weight function has property W2. These results yield no convergence orders (hence called order-less) but show that the Exact smoother always converges to the Approximate smoother uniformly pointwise and normwise as the density of the set X' increases. They will be applicable to Approximate smoothers generated by regular rectangular grids and sparse grids.

We then derive pointwise order of convergence results which are analogous to the results derived for the minimal norm interpolant in Subsections 2.5.2 and 2.5.3, and for the Exact smoother in Sections 4.6 and 4.7. Type 1 and Type 2 results are first derived for the case where no assumption is made regarding the unisolvency (Definition 66) of the independent data. Then it is shown that if the data is unisolvent of order m then the order of convergence of the smoother is also m.

5.6.1 Order-less convergence results

The results of this subsection only assume that the weight function has property W2 so that the data functions are continuous. Uniform pointwise convergence results are derived but **no order of convergence** results are obtained. To start with we will need a definition of convergence for *sequences of independent data sets*.

Definition 184 ***Convergence of independent data sets***

1. *A sequence of independent data* $X_n = \left(x_n^{(i)}\right)_{i=1}^{N_n}$ *is said to converge to the independent data* $X = \left(x^{(i)}\right)_{i=1}^{N}$, *denoted* $X_n \to X$, *if there exists an integer* K *such that* $N_n = N$ *when* $n \geq K$, *and for each* i, $\left|x_n^{(i)} - x^{(i)}\right| \to 0$ *as* $n \to \infty$.

2. *When* $k \geq K$, X_n *and* X *can be regarded as members of* $\mathbb{R}^{Nd}$ *and convergence as convergence in* $\mathbb{R}^{Nd}$ *under the Euclidean norm.*

Theorem 185 *Suppose s_e is the Exact smoother generated by the data $[X;y]$. Suppose X'_n is a sequence of independent data sets which converge to X in the sense of Definition 184 and that $s_a^{(n)}$ is the Approximate smoother generated by X'_n and $[X;y]$.*

Then the Approximate smoothers satisfy $J_e\left[s_a^{(n)}\right] \to J_e[s_e]$ as $n \to \infty$, where

$$J_e[f] = \rho \|f\|_{w,0}^2 + \frac{1}{N}|f_X - y|^2,$$

is the Exact smoother functional 4.1.

Proof. We first note that the definition of the convergence of independent data sets allows us to assume that the X'_n have the same number of points as X.

Now suppose X' is an arbitrary independent data set with the same number of points as X, and $s_a = s_a(X')$ is the corresponding Approximate smoother. If it can be shown that as a function of X', $J_e[s_a(X')]$ is continuous everywhere the the theorem holds since $J_e[s_e] = J_e[s_a(X)]$.

In fact, by Theorem 180

$$s_a(X')(x) = \sum_{i=1}^{N} R(x - x'_i)\,\alpha'_i,$$

where $\alpha' = (\alpha'_i)$ satisfies the matrix equation

$$\Psi\alpha' = R_{X',X}\,y,$$

and

$$\Psi = N\rho R_{X',X'} + R_{X,X'}^T R_{X,X'}.$$

Starting with part 3 of Corollary 177 and noting that $(\alpha')^T = \left(\Psi^{-1}R_{X',X}y\right)^T = y^T\overline{R}_{X,X'}\overline{\Psi}^{-1}$ we have

$$\begin{aligned}
J_e[s_a(X')] = \frac{1}{N}\left(y - s_a(X')_X\right)^T\overline{y} &= \frac{1}{N}|y|^2 - \frac{1}{N}s_a(X')_X^T\,\overline{y} \\
&= \frac{1}{N}|y|^2 - \frac{1}{N}\left(R_{X,X'}\alpha'\right)^T\overline{y} \\
&= \frac{1}{N}|y|^2 - \frac{1}{N}(\alpha')^T\left(R_{X,X'}\right)^T\overline{y} \\
&= \frac{1}{N}|y|^2 - \frac{1}{N}y^T R_{X,X'}\Psi^{-1}R_{X',X}\overline{y}.
\end{aligned}$$

If we can show that $R_{X,X'}$ and $\Psi_{X'}^{-1}$ are continuous functions of X' in a neighborhood of X, then we have $J_e[s_a(X')]$ is a continuous function of X'. But $R_{X',X'}$ and $R_{X,X'}$ are clearly continuous for all X' so $\Psi_{X'}$ and $\det\Psi_{X'}$ are continuous for all X'. Further, since $\Psi_{X'}$ is positive definite and regular for all X', $\det\Psi_{X'} > 0$ for all X' and it is clear from Crammer's rule that $\Psi_{X'}^{-1}$ is continuous everywhere. Thus $J_e[s_a(X')]$ is continuous everywhere and the proof is complete. ∎

The next corollary shows that the Approximate smoother converges to the Exact smoother, both pointwise and norm-wise, as the grid size goes to zero.

Corollary 186 *Suppose s_e is the Exact smoother generated by the data $[X;y]$ and that we have a sequence of regular grids X'_n with a common rectangular grid boundary and grid sizes h_n. Suppose X lies in the interior of the common grid boundary. Also, let $s_a(X'_n)$ denote the Approximate smoother generated by X'_n and $[X;y]$.*

Then $|h_n| \to 0$ implies $\|s_a(X'_n) - s_e\|_{w,0} \to 0$ and $\|s_a(X'_n) - s_e\|_\infty \to 0$, where the supremum norm $\|\cdot\|_\infty$ is defined on $\mathbb{R}^d$.

Proof. For each data point $x^{(k)} \in X$ there exists a sequence of distinct points $\left(z_n^{(k)}\right)_{n=1}^{\infty}$ such that $z_n^{(k)} \in X'_n$ and $z_n^{(k)} \to x^{(k)}$ in $\mathbb{R}^d$ as $n \to \infty$.

Set $Z_n = \left\{z_n^{(k)}\right\}_{k=1}^{N}$ and let $s_a(Z_n)$ be the Approximate smoother generated by Z_n. Then $Z_n \to X$ as independent data and by Theorem 185, $J_e[s_a(Z_n)] \to J_e[s_e]$.

But since $Z_n \subset X'_n$, we have $J_e[s_a(X'_n)] \le J_e[s_a(Z_n)]$. Also from the definition of s_e we have $J_e[s_e] \le J_e[s_a(X'_n)]$. Thus

$$J_e[s_e] \le J_e[s_a(X'_n)] \le J_e[s_a(Z_n)],$$

and so $J_e[s_a(X'_n)] \to J_e[s_e]$. To establish the convergence of $s_a(X'_n)$ we use part 1 of Theorem 4.12 which can be written

$$J_e[s_e] + \rho\left\|s_e - f\right\|_{w,0}^2 + \frac{1}{N}\sum_{k=1}^{N}\left|s_e\left(x^{(k)}\right) - f\left(x^{(k)}\right)\right|^2 = J_e[f],$$

for all $f \in X_w^0$. Choosing $f = s_a(X'_n)$ we see that $\left\|s_e - s_a(X'_n)\right\|_{w,0} \to 0$.

Finally, if R_x is the Riesz representer of the functional $f \to f(x)$

$$\begin{aligned}\left|s_e(x) - s_a(X'_n)(x)\right| = \left|(s_e - s_a(X'_n), R_x)_{w,0}\right| &\le \left\|s_e - s_a(X'_n)\right\|_{w,0}\left\|R_x\right\|_{w,0} \\ &= \left\|s_e - s_a(X'_n)\right\|_{w,0}\sqrt{R_0(0)},\end{aligned}$$

with the right side independent of x. ■

We can remove the constraint that the sets X'_n are regular, regular grids:

Corollary 187 *Suppose Ω is a bounded, open connected set. Suppose s_e is the Exact smoother generated by the data $[X;y]$ and that $X \subset \Omega$. Suppose also that we have a sequence of sets $X'_n \subset \Omega$ such that $\max\limits_{x\in\Omega}\operatorname{dist}(x, X'_n) \to 0$. Finally let $s_a(X'_n)$ denote the Approximate smoother generated by X'_n and $[X;y]$.*

Then $\left\|s_a(X'_n) - s_e\right\|_{w,0} \to 0$ and $\left\|s_a(X'_n) - s_e\right\|_\infty \to 0$, where the supremum norm $\left\|\cdot\right\|_\infty$ is defined on $\mathbb{R}^d$.

Proof. Let $X = \left\{x^{(k)}\right\}_{k=1}^N$. Since $X \subset \Omega$ the assumption that $\max\limits_{x\in\Omega}\operatorname{dist}(x, X'_n) \to 0$ implies $\operatorname{dist}\left(x^{(k)}, X'_n\right) \to 0$ and we can now use the arguments of the proof of the previous corollary. ■

5.6.2 General error results

The results of this subsection only assume that the weight function has property W2. The next result establishes some upper bounds for the pointwise difference between the Exact and Approximate smoothers which are uniform on $\mathbb{R}^d$. Here ρ is called the smoothing coefficient and is used to define the Exact smoother functional. No data densities are involved.

Theorem 188 *If s_e is the Exact smoother of the data $\left[X;\widetilde{\mathcal{E}}_X f_d\right]$ and s_a is the Approximate smoother generated by $\left[X;\widetilde{\mathcal{E}}_X f_d\right]$ and X' then*

$$\left|s_e(x) - s_a(x)\right| \le \left\|s_e\right\|_{w,0}\sqrt{R_0(0)}, \quad x \in \mathbb{R}^d, \tag{5.16}$$

and

$$\left|\mathcal{S}_X^e R_y(x) - \mathcal{S}_{X,X'}^a R_y(x)\right| \le R_0(0)\min\left\{1, \frac{R_0(0)}{\rho}\right\}, \quad x, y \in \mathbb{R}^d, \tag{5.17}$$

and $\left\|s_e\right\|_{w,0} \le \left\|f_d\right\|_{w,0}$.

Proof. From Corollary 176 we have $s_a = \mathcal{I}_{X'}s_e$ so that

$$\begin{aligned}\left|s_e(x) - s_a(x)\right| \le \left|s_e(x) - (\mathcal{I}_{X'}s_e)(x)\right| = \left|(s_e - \mathcal{I}_{X'}s_e, R_x)_{w,0}\right| &\le \left\|s_e - \mathcal{I}_{X'}s_e\right\|_{w,0}\left\|R_x\right\|_{w,0} \\ &\le \left\|s_e\right\|_{w,0}\left\|R_x\right\|_{w,0} \\ &= \left\|s_e\right\|_{w,0}\sqrt{R_0(0)},\end{aligned}$$

where the last equality follows directly from the definition of R_x. Substituting $f_d = R_y$ into 5.16 we get

$$\left|\mathcal{S}_X^e R_y(x) - \mathcal{S}_{X,X'}^a R_y(x)\right| \le \left\|\mathcal{S}_X^e R_y\right\|_{w,0}\sqrt{R_0(0)}, \quad x, y \in \mathbb{R}^d.$$

From equation 4.15, $\|\mathcal{S}_X^e f\|_{w,0} \leq \|f\|_{w,0} \sqrt{R_0(0)} \min\left\{1, \frac{R_0(0)}{\rho}\right\}$, so that

$$\|\mathcal{S}_X^e R_y\|_{w,0} \leq \|R_y\|_{w,0} \|R_0\|_{w,0} \min\left\{1, \frac{R_0(0)}{\rho}\right\} = R_0(0) \min\left\{1, \frac{R_0(0)}{\rho}\right\}.$$

and hence

$$\left|\mathcal{S}_X^e R_y(x) - \mathcal{S}_{X,X'}^a R_y(x)\right| \leq R_0(0) \min\left\{1, \frac{R_0(0)}{\rho}\right\}.$$

That $\|s_e\|_{w,0} \leq \|f_d\|_{w,0}$ follows from part 2 of Summary 174. ■

5.6.3 *Error estimates derived without assuming unisolvent data sets*

We will now derive order estimates for the pointwise difference between the Approximate smoother and the Exact smoother. The derivation of these estimates will rely strongly on the fact that the Approximate smoother is the interpolant of the Exact smoother on the set X' (see 5.3) and so we will use the convergence results for the minimal norm interpolant derived in Chapter 4. The convergence results for the interpolant and Exact smoother all involved one of three assumptions concerning the unisolvency of the independent data set X and the weight function w and its parameter κ:

1. w has property W2 for some $\kappa \geq 0$ with no assumption that X is unisolvent;
2. $\kappa \geq 1$ with no assumption that X is unisolvent;
3. $\kappa \geq 1$ and X is unisolvent.

We will consider the first two cases in this subsection.

Type 1 estimates ($\kappa \geq 0$)

The next result is an estimate of the difference between the Exact and Approximate smoothers for an arbitrary data function in X_w^0.

Theorem 189 *Suppose:*

1. *The weight function w has property W2 and that G is the basis function generated by w. Assume that for some s, C_G, $h_G > 0$ the basis function satisfies*
$$|G(0) - G(x)| \leq C_G |x|^{2s}, \quad |x| \leq h_G. \tag{5.18}$$
2. *Let s_e be the Exact smoother generated by the data $\left[X; \widetilde{\mathcal{E}}_X f_d\right]$.*
3. *Suppose that s_a is the Approximate smoother generated by the data $\left[X; \widetilde{\mathcal{E}}_X f_d\right]$ and the points X' contained in K, where K is a closed, bounded, infinite set.*
4. *Let $\mathcal{I}_{X'}$ be the minimal norm interpolant operator on the set X'.*

Then if $k_G = (2\pi)^{-\frac{d}{4}} \sqrt{2C_G}$ we have the error bound

$$|s_e(x) - s_a(x)| \leq \sqrt{(s_e - s_a, s_e)_{w,0}}\, k_G (h_{X',K})^s, \quad x \in K, \tag{5.19}$$

when $h_{X',K} = \max\limits_{x \in K} \operatorname{dist}(x, X') \leq h_G$.

Further, $\sqrt{(s_e - s_a, s_e)_{w,0}} \leq \|s_e\|_{w,0} \leq \|f_d\|_{w,0}$ and the order of convergence is at least s in $h_{X',K}$.

Proof. From Corollary 179 we have $s_a = \mathcal{I}_{X'} s_e$. Now we can apply Theorem 59 which estimates the error of the minimal norm interpolant of an arbitrary data function. In this case the data function is s_e, the data points are X', the data region is K, and so for $x \in K$

$$\begin{aligned} |s_e(x) - s_a(x)| = |s_e(x) - (\mathcal{I}_{X'} s_e)(x)| &\leq \sqrt{(s_e - \mathcal{I}_{X'} s_e, s_e)_{w,0}}\, k_G (h_{X',K})^s \\ &= \sqrt{(s_e - s_a, s_e)_{w,0}}\, k_G (h_{X',K})^s, \end{aligned}$$

when $h_{X',K} < h_G$. Finally, from part 2 of Summary 174, $I - \mathcal{I}_{X'}$ and $\mathcal{S}_X^e$ are contractions and so

$$\sqrt{(s_e - \mathcal{I}_{X'} s_e, s_e)_{w,0}} \leq \sqrt{\|s_e - \mathcal{I}_{X'} s_e\|_{w,0} \|s_e\|_{w,0}} \leq \|s_e\|_{w,0} \leq \|f_d\|_{w,0} .$$

■

Remark 190

1. *Unlike the error estimates for the Approximate smoother and the minimal norm interpolant, this estimate for* $|s_e(x) - s_a(x)|$ *does not require the independent data set* X *to be in a closed, bounded, infinite set. Instead, it is* X' *which is required to be in such a set. The density of* X *is not explicitly involved in the estimate i.e. there is no term involving* $h_{X,K}$ *and the dependency on* X *is in the formula for* s_e *and can thus be eliminated by the approximation* $\sqrt{(s_e - \mathcal{I}_{X'} s_e, s_e)_{w,0}} \leq \|f_d\|_{w,0}$.

2. *This result confirms the convergence result of Corollary 186. Unlike the Approximate smoother error estimates, this estimate does not explicitly involve the smoothing coefficient* ρ *and has the same form as the interpolation error estimates. However, the smoothing coefficient is part of the formula for* s_e *and can be eliminated by the approximation* $\sqrt{(s_e - \mathcal{I}_{X'} s_e, s_e)_{w,0}} \leq \|f_d\|_{w,0}$.

3. *The following argument shows the factor* $\sqrt{(s_e - s_a, s_e)_{w,0}}$ *in the estimate 5.19 can be calculated numerically. Indeed, since* $s_a = \mathcal{I}_{X'} s_e$ *we have* $(s_e - s_a, s_a)_{w,0} = (s_e - s_a, \mathcal{I}_{X'} s_e)_{w,0} = (\mathcal{I}_{X'} s_e - \mathcal{I}_{X'} s_a, s_e)_{w,0} = 0$ *so that* $(s_e, s_a)_{w,0} = \|s_a\|_{w,0}^2$ *and*

   $(s_e - s_a, s_e)_{w,0} = \sqrt{\|s_e\|_{w,0}^2 - \|s_a\|_{w,0}^2}$. *Part 2 of Corollary 177 and part 3 of Theorem 145 can then be used to calculate* $\|s_a\|_{w,0}$ *and* $\|s_e\|_{w,0}$ *respectively.*

As with the Exact smoother error estimates which were valid for arbitrary data functions in X_w^0 we can try to improve the last convergence result for data functions which have the form R_y, $y \in K$.

Theorem 191 *Suppose we have the same assumptions and notation as Theorem 189 except that now the data function is specialized to* R_y. *Hence* $\mathcal{S}_X^e R_y$ *is the Exact smoother and* $\mathcal{S}_{X,X'}^a R_y$ *is the Approximate smoother.*

Further, suppose that the data region K *that contains* X *is a bounded, closed and infinite set.*

Then if $h_{X,K} \leq h_G$ *and* $h_{X',K} \leq h_G$ *we have the estimate*

$$\left|\mathcal{S}_X^e R_y(x) - \mathcal{S}_{X,X'}^a R_y(x)\right| \leq k_G (h_{X',K})^s \left(\sqrt{\rho N} + k_G (h_{X,K})^s + k_G (h_{X',K})^s\right), \quad x, y \in K, \tag{5.20}$$

where $k_G = (2\pi)^{-\frac{d}{4}} \sqrt{2C_G}$ *and* C_G *satisfies 5.18.*

Proof. Theorem 189 with $f_d = R_y$ yields the estimate

$$\left|\mathcal{S}_X^e R_y(x) - \mathcal{S}_{X,X'}^a R_y(x)\right| \leq k_G \sqrt{(\mathcal{S}_X^e R_y - \mathcal{I}_{X'} \mathcal{S}_X^e R_y, \mathcal{S}_X^e R_y)_{w,0}} (h_{X',K})^s , \tag{5.21}$$

when $h_{X',K} \leq h_G$ and $x \in K$. The term $\sqrt{(\mathcal{S}_X^e R_y - \mathcal{I}_{X'} \mathcal{S}_X^e R_y, \mathcal{S}_X^e R_y)_{w,0}}$ will now be manipulated in order to try to improve the convergence. Since $\mathcal{I}_{X'}$ is a self-adjoint projection we have

$$\begin{aligned}
(\mathcal{S}_X^e R_y - \mathcal{I}_{X'} \mathcal{S}_X^e R_y, \mathcal{S}_X^e R_y)_{w,0} &= ((I - \mathcal{I}_{X'}) \mathcal{S}_X^e R_y, \mathcal{S}_X^e R_y)_{w,0} \qquad (5.22)\\
&= ((I - \mathcal{I}_{X'}) \mathcal{S}_X^e R_y, R_y - (I - \mathcal{S}_X^e) R_y)_{w,0} \\
&= ((I - \mathcal{I}_{X'}) \mathcal{S}_X^e R_y, R_y)_{w,0} - ((I - \mathcal{I}_{X'}) \mathcal{S}_X^e R_y, (I - \mathcal{S}_X^e) R_y)_{w,0} \\
&= ((I - \mathcal{I}_{X'}) \mathcal{S}_X^e R_y)(y) - ((I - \mathcal{S}_X^e)(I - \mathcal{I}_{X'}) \mathcal{S}_X^e R_y, R_y)_{w,0} \\
&= ((I - \mathcal{I}_{X'}) \mathcal{S}_X^e R_y)(y) - ((I - \mathcal{S}_X^e)(I - \mathcal{I}_{X'}) \mathcal{S}_X^e R_y)(y) \\
&\leq |((I - \mathcal{I}_{X'}) \mathcal{S}_X^e R_y)(y)| + |((I - \mathcal{S}_X^e)(I - \mathcal{I}_{X'}) \mathcal{S}_X^e R_y)(y)| . \qquad (5.23)
\end{aligned}$$

The application of the Type 1 interpolation convergence estimate 3.55 to the data function $\mathcal{S}_X^e R_y$ gives

$$|((I - \mathcal{I}_{X'}) \mathcal{S}_X^e R_y)(y)| \leq ((I - \mathcal{I}_{X'}) \mathcal{S}_X^e R_y, \mathcal{S}_X^e R_y)_{w,0} \, k_G \left(h_{X',K}\right)^s, \tag{5.24}$$

when $h_{X',K} < h_G$ and $y \in K$. Further, for an arbitrary data function $f_d \in X_w^0$, the Exact smoother estimate 4.22 implies

$$|((I - \mathcal{S}_X^e) f_d)(x)| \leq \sqrt{(f_d - \mathcal{S}_X^e f_d, f_d)_{w,0}} \left(\sqrt{\rho N} + k_G \left(h_{X,K}\right)^s\right), \quad x \in K,$$

when $h_{X,K} \leq h_G$. But in this case $f_d = (I - \mathcal{I}_{X'}) \mathcal{S}_X^e R_y$ where $x, y \in K$, and so

$$|((I - \mathcal{S}_X^e)(I - \mathcal{I}_{X'}) \mathcal{S}_X^e R_y)(x)| \leq \sqrt{(f_d - \mathcal{S}_X^e f_d, f_d)_{w,0}} \left(\sqrt{\rho N} + k_G \left(h_{X,K}\right)^s\right),$$

and

$$\sqrt{(f_d - \mathcal{S}_X^e f_d, f_d)_{w,0}} \leq \sqrt{(f_d, f_d)_{w,0}} = \sqrt{((I - \mathcal{I}_{X'}) \mathcal{S}_X^e R_y, \mathcal{S}_X^e R_y)_{w,0}},$$

so that

$$\begin{aligned} |((I - \mathcal{S}_X^e)(I - \mathcal{I}_{X'}) \mathcal{S}_X^e R_y)(x)| \leq &\sqrt{((I - \mathcal{I}_{X'}) \mathcal{S}_X^e R_y, \mathcal{S}_X^e R_y)_{w,0}} \times \\ &\times \left(\sqrt{\rho N} + k_G \left(h_{X,K}\right)^s\right). \end{aligned} \tag{5.25}$$

Next using 5.24 and 5.25 to estimate the right side of 5.23 we obtain

$$\begin{aligned} ((I - \mathcal{I}_{X'}) \mathcal{S}_X^e R_y, \mathcal{S}_X^e R_y)_{w,0} \leq &\sqrt{((I - \mathcal{I}_{X'}) \mathcal{S}_X^e R_y, \mathcal{S}_X^e R_y)_{w,0}} k_G \left(h_{X',K}\right)^s + \\ &+ \sqrt{((I - \mathcal{I}_{X'}) \mathcal{S}_X^e R_y, \mathcal{S}_X^e R_y)_{w,0}} \left(\sqrt{\rho N} + k_G \left(h_{X,K}\right)^s\right), \end{aligned}$$

so that

$$\sqrt{((I - \mathcal{I}_{X'}) \mathcal{S}_X^e R_y, \mathcal{S}_X^e R_y)_{w,0}} \leq \sqrt{\rho N} + k_G \left(h_{X,K}\right)^s + k_G \left(h_{X',K}\right)^s,$$

and as a consequence of equation 5.22, inequality 5.21 implies

$$\begin{aligned} \left|\mathcal{S}_X^e R_y(x) - \mathcal{S}_{X,X'}^a R_y(x)\right| &\leq k_G \left(h_{X',K}\right)^s \sqrt{((I - \mathcal{I}_{X'}) \mathcal{S}_X^e R_y, \mathcal{S}_X^e R_y)_{w,0}} \\ &\leq k_G \left(h_{X',K}\right)^s \left(\sqrt{\rho N} + k_G \left(h_{X,K}\right)^s + k_G \left(h_{X',K}\right)^s\right), \end{aligned}$$

as claimed. ■

Remark 192 *Our attempt at improving convergence has failed and the order of convergence in $h_{X',K}$ is still s. Unlike the cases of the interpolation error in Subsection 2.5.2 and the Exact smoother error of Section 4.6 this attempt to improve the estimate for the order of convergence of the Approximate smoother to the Exact smoother has resulted in new terms appearing. Essentially the original term $k_G \left(h_{X,K}\right)^s$ has been squared but the term $k_G \left(h_{X',K}\right)^s \left(\sqrt{\rho N} + k_G \left(h_{X,K}\right)^s\right)$ has been added. Hence, to get the benefit of squaring, the term $\sqrt{\rho N} + k_G \left(h_{X,K}\right)^s$ should be the much the same size as $k_G \left(h_{X',K}\right)^s$.*

The next result gives some idea of how the Exact and Approximate smoothers compare when the smoothing parameter ρ is large.

Theorem 193 *Suppose the assumptions and notation of Theorem 189 hold so that s_e is the Exact smoother and s_a is the Approximate smoother of the data function f_d. Then for $\rho > 0$*

$$|s_e(x) - s_a(x)| \leq \sqrt{(s_e - f_d, f_d)_{w,0}} \left(k_G\right)^2 \frac{\left(h_{X',K}\right)^{2s}}{\sqrt{\rho}}, \quad x \in K, \; f_d \in X_w^0, \tag{5.26}$$

with $\sqrt{(s_e - f_d, f_d)_{w,0}} \leq \|f_d\|_{w,0}$. For given ρ the order of convergence is $2s$.

Proof. From part 4 of Summary 174

$$s_e = \widetilde{\mathcal{E}}_X^* \left(N\rho I + R_{X,X} \right)^{-1} \widetilde{\mathcal{E}}_X f_d,$$

so that

$$\begin{aligned}
\left(s_e - s_a, s_e \right)_{w,0} &= \left(s_e - s_a, \widetilde{\mathcal{E}}_X^* \left(N\rho I + R_{X,X} \right)^{-1} \widetilde{\mathcal{E}}_X f_d \right)_{w,0} \\
&= \left(\widetilde{\mathcal{E}}_X \left(s_e - s_a \right), \left(N\rho I + R_{X,X} \right)^{-1} \widetilde{\mathcal{E}}_X f_d \right) \\
&\leq \left| \widetilde{\mathcal{E}}_X \left(s_e - s_a \right) \right| \left| \left(N\rho I + R_{X,X} \right)^{-1} \widetilde{\mathcal{E}}_X f_d \right|
\end{aligned}$$

But from part 2 Lemma 142

$$\left\| s_e - f_d \right\|_{w,0}^2 = \left(s_e - f_d, f_d \right)_{w,0} - \rho N \left| \left(\rho N I + R_{X,X} \right)^{-1} \widetilde{\mathcal{E}}_X f_d \right|^2,$$

so that $\left| \left(\rho N I + R_{X,X} \right)^{-1} \widetilde{\mathcal{E}}_X f_d \right| \leq \left(\rho N \right)^{-1/2} \sqrt{\left(s_e - f_d, f_d \right)_{w,0}}$.
Hence, if $\left| s_e - s_a \right|_{\infty,K} = \max\limits_{x \in K} \left| s_e(x) - s_a(x) \right|$,

$$\begin{aligned}
\left(s_e - s_a, s_e \right)_{w,0} &\leq \left| \widetilde{\mathcal{E}}_X \left(s_e - s_a \right) \right| \sqrt{\left(s_e - f_d, f_d \right)_{w,0}} \left(\rho N \right)^{-1/2} \\
&\leq \sqrt{N} \left| s_e - s_a \right|_{\infty,K} \sqrt{\left(s_e - f_d, f_d \right)_{w,0}} \left(\rho N \right)^{-1/2} \\
&= \left| s_e - s_a \right|_{\infty,K} \sqrt{\left(s_e - f_d, f_d \right)_{w,0}} \rho^{-1/2},
\end{aligned} \tag{5.27}$$

and by 5.19

$$\left| s_e - s_a \right|_{\infty,K} \leq \sqrt{\left(s_e - s_a, s_e \right)_{w,0}} k_G \left(h_{X',K} \right)^s,$$

which implies

$$\begin{aligned}
\left| s_e - s_a \right|_{\infty,K} &\leq \sqrt{\left| s_e - s_a \right|_{\infty,K} \sqrt{\left(s_e - f_d, f_d \right)_{w,0}} \rho^{-1/2} \left(k_G \left(h_{X',K} \right)^s \right)} \\
&= \sqrt{\left| s_e - s_a \right|_{\infty,K}} \sqrt[4]{\left(s_e - f_d, f_d \right)_{w,0}} \rho^{-1/4} k_G \left(h_{X',K} \right)^s,
\end{aligned}$$

and hence

$$\left| s_e - s_a \right|_{\infty,K} \leq \sqrt{\left(s_e - f_d, f_d \right)_{w,0}} \left(k_G \right)^2 \frac{\left(h_{X',K} \right)^{2s}}{\sqrt{\rho}},$$

which implies 5.26. ■

Remark 194 *Theorem 193 suggests that as the smoothing parameter ρ increases the Approximate smoother converges to the Exact smoother. Further, for any 'large' value of ρ the rate of convergence in $h_{X',K}$ is at least $\left(h_{X',K} \right)^{2s}$. Of course, as the smoothing parameter increases the Exact smoother gets further from the data function.*

The next result summarizes the estimates of $\left| s_e(x) - s_a(x) \right|$ and $\left| \mathcal{S}_X^e R_y(x) - \mathcal{S}_{X,X'}^a R_y(x) \right|$ we have derived above.

Corollary 195 *For an arbitrary data function f_d we have the error estimates*

$$\left| s_e(x) - s_a(x) \right| \leq \min \left\{ \begin{array}{l} \left\| f_d \right\|_{w,0} k_G \left(h_{X',K} \right)^s, \\ \left\| f_d \right\|_{w,0} \sqrt{R_0(0)}, \\ \left\| f_d \right\|_{w,0} \left(k_G \right)^2 \frac{\left(h_{X',K} \right)^{2s}}{\sqrt{\rho}}. \end{array} \right. , \quad x \in K, \tag{5.28}$$

and for the Riesz data functions R_y we have the error estimates:

$$\left|\mathcal{S}_X^e R_y(x) - \mathcal{S}_{X,X'}^a R_y(x)\right| \leq \min \begin{cases} k_G \left(h_{X',K}\right)^s \left(\sqrt{\rho N} + k_G \left(h_{X,K}\right)^s + k_G \left(h_{X',K}\right)^s\right), \\ R_0(0) \min\left\{1, \frac{R_0(0)}{\rho}\right\}, \\ \left(k_G\right)^2 \frac{\left(h_{X',K}\right)^{2s}}{\sqrt{\rho}} \left(\sqrt{\rho N} + k_G \left(h_{X,K}\right)^s\right), \\ \left(k_G\right)^2 \frac{\left(h_{X',K}\right)^{2s}}{\sqrt{\rho}} \sqrt{R_0(0)}, \\ \left(k_G\right)^2 \frac{\left(h_{X',K}\right)^{2s}}{\sqrt{\rho}} \sqrt{\left\|R_y\right\|_{\infty,K} + R_0(0) \min\left\{1, \frac{R_0(0)}{\rho}\right\}}. \end{cases} \tag{5.29}$$

for $x, y \in K$.

Proof. The inequalities 5.28 are the estimates 5.16, 5.19 and 5.26 simplified using the inequalities $\sqrt{(s_e - s_a, s_e)_{w,0}} \leq \|f_d\|_{w,0}$ and $\sqrt{(s_e - f_d, f_d)_{w,0}} \leq \|f_d\|_{w,0}$.

Regarding the inequalities 5.29: The first inequality is 5.20 and the second inequality is 5.17. When $f_d = R_y$ is substituted into 5.26 we get for $x \in K$

$$\left|\mathcal{S}_X^e R_y(x) - \mathcal{S}_{X,X'}^a R_y(x)\right| \leq \left(k_G\right)^2 \frac{\left(h_{X',K}\right)^{2s}}{\sqrt{\rho}} \sqrt{\mathcal{S}_X^e R_y(x) - R_y(x)},$$

and then substituting the estimates 5.42 for $\mathcal{S}_X^e R_y(x) - R_y(x)$ immediately gives the remaining three inequalities. ■

Type 1 examples

The weight function examples used here are those used in the interpolation and Exact smoother documents i.e. the radial shifted thin-plate splines, Gaussian and Sobolev splines, and the tensor product extended B-spline weight functions, augmented by the tensor product central difference weight functions from Chapter 3.

When the smoothing coefficient ρ is zero the error estimates of this subsection are the Type 1, non-unisolvent, interpolation estimates of Subsubsection 2.5.2 and consequently the formulas or values for s, C_G, h_G, k_G have already been calculated and were given in Table 2.1. We give these values here in Table 5.1 together with the Type 1 values from Table 3.1 for the central difference tensor product weight functions.

Smoother convergence order estimates				
Type 1 non-unisolvent estimates: $\kappa \geq 0$. $k_G = (2\pi)^{-d/4} \sqrt{2C_G}$.				
Weight function	Parameter constraints	Converg. order s	C_G	h_G
Sobolev splines	$v - \frac{d}{2} = 1$	$\frac{1}{2}$	$\frac{\left\|\rho K_0(\rho)\right\|_\infty}{2^{v-1}\Gamma(v)}$ (2)	∞
$(v > d/2)$	$v - \frac{d}{2} \neq 1$	1	$\frac{\left\|D^2 \widetilde{K}_{v-d/2}\right\|_\infty}{2^v \Gamma(v)}$ (2)	”
Shifted thin-plate $(-d/2 < v < 0)$	-	1	eq. (5.30)	∞
Gaussian	-	1	$2e^{-3/2}$	∞
Extended B-spline	-	$\frac{1}{2}$	$G_1(0)^{d-1} \left\|DG_1\right\|_\infty \sqrt{d}$ (1)	∞
Central difference	-	$\frac{1}{2}$	$G_1(0)^{d-1} \left\|DG_1\right\|_\infty \sqrt{d}$ (1)	∞
(1) G_1 is the univariate basis function used to form the tensor product				
(2) K_v is the modified Bessel function and $\widetilde{K}_v(r) = r^v K_v(r)$.				

TABLE 5.1.

$$C_G = \left|\left(2rf'' + f'\right)(r_{\max})\right|, \quad \textit{where } f(r) = (1+r)^v \textit{ and } r_{\max} = \frac{1-2v}{3}. \tag{5.30}$$

Type 2 estimates ($\kappa \geq 1$)

These results can be proved by first using the theorems of Subsection 4.6.3 to obtain estimates of the form $|G(0) - G(x)| \leq C_G |x|^{2s}$ and then employing Corollary 195.

Theorem 196 *Suppose a weight function satisfies property W2 for some $\kappa \geq 1$ and denote the basis function by G. Then the smoother error estimates of Corollary 195 hold for*

$$C_G = -\frac{1}{2}\left(\nabla^2 G\right)(0)\, d, \quad s = 1, \quad h_G = \infty, \quad k_G = (2\pi)^{-\frac{d}{4}} \sqrt{-\left(\nabla^2 G\right)(0)}\sqrt{d}. \tag{5.31}$$

However, if the weight function is radial we can use the estimates of:

Theorem 197 *Suppose a radial weight function satisfies property W2 for $\kappa = 1$ and denote the (radial) basis function by G. Set $r = |x|$. Then:*

1. *If $G(x) = f\left(r^2\right)$ the smoother error estimates of Corollary 195 hold for*

$$C_G = -f'(0)\, d^2, \quad s = 1, \quad h_G = \infty, \quad k_G = (2\pi)^{-\frac{d}{4}} \sqrt{-2f'(0)}d.$$

2. *If $G(x) = g(r)$ then the smoother error estimates of Corollary 195 hold for*

$$C_G = -\frac{1}{2}g''(0)\, d^2, \quad s = 1, \quad h_G = \infty, \quad k_G = (2\pi)^{-\frac{d}{4}} \sqrt{-g''(0)}d. \tag{5.32}$$

If the weight function is a tensor product the following result will be useful:

Theorem 198 *Suppose a tensor product weight function satisfies property W2 for $\kappa = 1$ and denote the univariate basis function by G_1. Then the smoother error estimates of Corollary 195 hold when*

$$C_G = -\frac{d}{2}G_1(0)^{d-1} D^2 G_1(0), \quad s = 1, \quad h_G = \infty, \quad k_G = (2\pi)^{-\frac{d}{4}} \sqrt{-G_1(0)^{d-1} D^2 G_1(0)}\sqrt{d}.$$

Type 2 examples

When the smoothing coefficient is zero the smoother error estimates of the Theorems of the previous subsubsection become algebraically identical to the Type 2 interpolant error estimates of Chapters 2 and 3. Further, the weight function examples used above were also used for the Type 2 interpolation examples of Chapter 2. If we use the 'radial' Theorem 197 to do the estimates for the radial basis functions and Theorem 196 to do the estimates for the tensor product basis functions then the values for s, C_G, h_G, k_G will match those obtained for the interpolants. These are given below in Table 5.2 which is Table 2.2 augmented by the results for the central difference tensor product weight functions from Table 3.1.

When $\rho = 0$ we see that the order of convergence is at least 1 for an arbitrary data function and at least 2 for a Riesz representer data function, no matter what value κ takes. However, in the next section we will show that by assuming the independent data is unisolvent of order $\kappa \geq 1$ it follows that an order of convergence of at least $\lfloor \kappa \rfloor$ can be attained for an arbitrary data function and an order of convergence of $\lfloor 2\kappa \rfloor$ for a Riesz representer data function.

5.6.4 Error estimates using unisolvent data sets

Unisolvent sets were introduced in Subsection 2.5.3 where they were used to define the Lagrange interpolation operators $\mathcal{P}$ and $\mathcal{Q} = I - \mathcal{P}$. Lemma 72 supplied the key interpolation theory and I reproduce it here:

Lemma 199 *(Copy of Lemma 72)Suppose first that:*

1. *Ω is a bounded, open, connected subset of $\mathbb{R}^d$ having the cone property.*
2. *X is a unisolvent subset of Ω of order κ.*

   Now define

$$h_{X,\Omega} = \sup_{\omega \in \Omega} \operatorname{dist}(\omega, X),$$

Smoother convergence order estimates			
Type 2 non-unisolvent estimates assuming $\kappa \geq 1$			
Weight function	Parameter constraints	Converg. order	$(2\pi)^{d/4}\, k_G/\sqrt{d}$
Sobolev splines	$v - \frac{d}{2} \geq 2$	1	$\sqrt{\frac{\Gamma(v-d/2-1)}{2^{d/2+1}\Gamma(v)}}$
$(v > d/2)$	$1 < v - \frac{d}{2} < 2$	1	$\sqrt{\frac{\Gamma(v-d/2-1)}{2^{2v-d/2-3}\Gamma(v)}}$
Shifted thin-plate $(-d/2 < v < 0)$	-	1	$\sqrt{-2v}$
Gaussian	-	1	$\sqrt{2}$
Extended B-spline $(1 \leq n \leq l)$	$n \geq 2$	1	$\sqrt{-G_1(0)^{d-1}\, D^2 G_1(0)}$ [(1)]
Central difference $(1 \leq n \leq l)$	$n \geq 2$	1	$\sqrt{-G_1(0)^{d-1}\, D^2 G_1(0)}$ [(1)]
[(1)] G_1 is the univariate basis function used to form the tensor product.			

TABLE 5.2.

and fix $x \in X$. By using Lagrange polynomial interpolation techniques it can be shown there are constants $c_{\Omega,\kappa}, h_{\Omega,\kappa} > 0$ such that when $h_{X,\Omega} < h_{\Omega,\kappa}$ there exists a minimal unisolvent set $A \subset X$ satisfying

$$\operatorname{diam}(A \cup \{x\}) \leq c_{\Omega,\kappa} h_{X,\Omega}.$$

Further, suppose $\{l_j\}_{j=1}^{M}$ is the cardinal basis of P_κ with respect to a minimal unisolvent subset of Ω. Again, using Lagrange interpolation techniques, it can be shown there exists a constant $K'_{\Omega,\kappa} > 0$ such that

$$\sum_{j=1}^{M} |l_j(x)| \leq K'_{\Omega,\kappa},$$

for all $x \in \Omega$ and all minimal unisolvent subsets of Ω.

To derive our estimates for $|s_e(x) - s_a(x)|$ we will rely on the formula $s_a = \mathcal{I}_{X'} s_e$ of part 2 Corollary 176. Hence we will also need the minimal norm interpolation convergence results of Subsection 2.5.3. We are now ready to state our order of convergence result for an arbitrary data function in X_w^0:

Theorem 200 *Suppose:*

1. *w is a weight function with property W2 for parameter $\kappa \geq 1$ and let G be the basis function generated by w. Set $m = \lfloor \kappa \rfloor$.*
2. *s_e is the Exact smoother generated by the data $\left[X; \widetilde{\mathcal{E}}_X f_d\right]$ for some data function $f_d \in X_w^0$.*
3. *s_a is the Approximate smoother generated by $\left[X; \widetilde{\mathcal{E}}_X f_d\right]$ and the set $X' \subset \Omega$.*
4. *$\mathcal{I}_{X'}$ is the minimal norm interpolant operator on the set X'.*
5. *We use the notation and assumptions of Lemma 199 which means that X' is κ-unisolvent and Ω is a bounded, open, connected set whose boundary satisfies the cone condition.*

Then there exist positive constants $k_G = \frac{d^{m/2}}{(2\pi)^{d/2}} (c_{\Omega,\kappa})^m K'_{\Omega,m} \max\limits_{|\beta|=m} \left|D^{2\beta} G(0)\right|$ and $h_{\Omega,m}$ such that

$$|s_e(x) - s_a(x)| \leq \sqrt{(s_e - s_a, s_e)_{w,0}}\; k_G \left(h_{X',\Omega}\right)^m, \quad x \in \overline{\Omega}, \tag{5.33}$$

when $h_{X',\Omega} = \sup\limits_{\omega' \in \Omega} \operatorname{dist}(\omega', X') < h_G$.
The constants $c_{\Omega,m}$, $K'_{\Omega,m}$ and $h_{\Omega,m}$ only depend on Ω, m and d. Further,
$\sqrt{(s_e - s_a, s_e)_{w,0}} \leq \|s_e\|_{w,0} \leq \|f_d\|_{w,0}$ and the order of convergence is at least $\lfloor \kappa \rfloor$ in $h_{X',\Omega}$.

Proof. From part 2 Corollary 176, $s_a = \mathcal{I}_{X'} s_e$ and by the interpolation estimate 2.50 there exists a constant k_G such that

$$\left| s_e(x) - (\mathcal{I}_{X'} s_e)(x) \right| \leq \sqrt{(s_e - \mathcal{I}_{X'} s_e, s_e)_{w,0}}\, k_G (h_{X',\Omega})^m, \quad x \in \overline{\Omega}, \tag{5.34}$$

when $h_{X',\Omega} < h_{\Omega,m}$. From Theorem 189, $\sqrt{(s_e - \mathcal{I}_{X'} s_e, s_e)_{w,0}} \leq \|s_e\|_{w,0} \leq \|f_d\|_{w,0}$ and so the order of convergence is at least m. ■

Remark 201 *The estimate derived in the last theorem has its dependency on the data points X contained in s_e in the expression $\sqrt{(s_e - s_a, s_e)_{w,0}}$.*

We now try to derive an improved convergence result from the last theorem for a data function which has the form R_y.

Theorem 202 *Suppose we have the same assumptions and notation as Theorem 200 except that now the data function is specialized to R_y, so that $\mathcal{S}^e_X R_y$ is the Exact smoother of R_y and $\mathcal{S}^a_{X,X'} R_y$ is the Approximate smoother of R_y.*

Further, suppose X is a unisolvent subset of Ω of order κ, and suppose that the data region Ω that contains X is a bounded, open, connected set which satisfies the cone condition. Set $m = \lfloor \kappa \rfloor$.

Then if $h_{X,K} \leq h_{\Omega,m}$ and $h_{X',K} \leq h_{\Omega,m}$ we have the estimate

$$\left| \mathcal{S}^e_X R_y(x) - \mathcal{S}^a_{X,X'} R_y(x) \right| \leq k_G (h_{X',\Omega})^m \left(K'_{\Omega,\kappa} \sqrt{\rho N} + k_G (h_{X,\Omega})^m + k_G (h_{X',\Omega})^m \right), \tag{5.35}$$

for all $x, y \in \overline{\Omega}$, where $k_G = \frac{d^{m/2}}{(2\pi)^{d/2}} (c_{\Omega,\kappa})^m K'_{\Omega,\kappa} \max\limits_{|\beta|=m} \left| D^{2\beta} G(0) \right|$.

The order of convergence is at least m in $h_{X',\Omega}$.

Proof. Theorem 200 with data function R_y implies

$$\left| \mathcal{S}^e_X R_y(x) - \mathcal{S}^a_{X,X'} R_y(x) \right| \leq k_G \sqrt{\left(\mathcal{S}^e_X R_y - \mathcal{S}^a_{X,X'} R_y, \mathcal{S}^e_X R_y \right)_{w,0}} (h_{X',\Omega})^m, \tag{5.36}$$

for $x \in \overline{\Omega}$, when $h_{X',\Omega} \leq h_{\Omega,m}$.

The term $\sqrt{\left(\mathcal{S}^e_X R_y - \mathcal{S}^a_{X,X'} R_y, \mathcal{S}^e_X R_y \right)_{w,0}}$ will now be manipulated to try to improve the convergence. Indeed, from inequality 5.23

$$\begin{aligned} \left(\mathcal{S}^e_X R_y - \mathcal{S}^a_{X,X'} R_y, \mathcal{S}^e_X R_y \right)_{w,0} &= \left((I - \mathcal{I}_{X'}) \mathcal{S}^e_X R_y, \mathcal{S}^e_X R_y \right)_{w,0} && (5.37) \\ &\leq \left| \left((I - \mathcal{I}_{X'}) \mathcal{S}^e_X R_y \right)(y) \right| + \\ &\quad + \left| \left((I - \mathcal{S}^e_X)(I - \mathcal{I}_{X'}) \mathcal{S}^e_X R_y \right)(y) \right|, && (5.38) \end{aligned}$$

and using 2.50 we estimate the interpolation error to be

$$\left| \left((I - \mathcal{I}_{X'}) \mathcal{S}^e_X R_y \right)(y) \right| \leq \sqrt{\left((I - \mathcal{I}_{X'}) \mathcal{S}^e_X R_y, \mathcal{S}^e_X R_y \right)_{w,0}}\, k_G (h_{X',\Omega})^m, \tag{5.39}$$

when $y \in \Omega$ and $h_{X',\Omega} \leq h_{\Omega,m}$, and using 4.31 we estimate the Exact smoothing error for an arbitrary data function $f_d \in X^0_w$ to be

$$\left| f_d(y) - (\mathcal{S}^e_X f_d)(y) \right| \leq \sqrt{(f_d - \mathcal{S}^e_X f_d, f_d)_{w,0}} \left(K'_{\Omega,\kappa} \sqrt{\rho N} + k_G (h_{X,\Omega})^m \right), \quad y \in \overline{\Omega}.$$

Consequently, when $f_d = (I - \mathcal{I}_{X'}) \mathcal{S}^e_X R_y$

$$\left| \left((I - \mathcal{S}^e_X)(I - \mathcal{I}_{X'}) \mathcal{S}^e_X R_y \right)(y) \right| \leq \sqrt{(f_d - \mathcal{S}^e_X f_d, f_d)_{w,0}} \left(K'_{\Omega,\kappa} \sqrt{\rho N} + k_G (h_{X,\Omega})^m \right),$$

and

$$\begin{aligned} \sqrt{(f_d - \mathcal{S}^e_X f_d, f_d)_{w,0}} \leq \sqrt{(f_d, f_d)_{w,0}} &= \|f_d\|_{w,0} = \left\| (I - \mathcal{I}_{X'}) \mathcal{S}^e_X R_y \right\|_{w,0} \\ &= \sqrt{\left((I - \mathcal{I}_{X'}) \mathcal{S}^e_X R_y, \mathcal{S}^e_X R_y \right)_{w,0}}, \end{aligned}$$

so that

$$\left|\left(\left(I-\mathcal{S}_X^e\right)\left(I-\mathcal{I}_{X'}\right)\mathcal{S}_X^e R_y\right)(y)\right| \leq \sqrt{\left(\left(I-\mathcal{I}_{X'}\right)\mathcal{S}_X^e R_y, \mathcal{S}_X^e R_y\right)_{w,0}}\left(K'_{\Omega,\kappa}\sqrt{\rho N}+k_G\left(h_{X,\Omega}\right)^m\right). \quad (5.40)$$

Next, using 5.39 and 5.40 to estimate the right side of 5.38 we obtain

$$\begin{aligned}\left(\left(I-\mathcal{I}_{X'}\right)\mathcal{S}_X^e R_y, \mathcal{S}_X^e R_y\right)_{w,0} &\leq \sqrt{\left(\left(I-\mathcal{I}_{X'}\right)\mathcal{S}_X^e R_y, \mathcal{S}_X^e R_y\right)_{w,0}}\left(K'_{\Omega,\kappa}\sqrt{\rho N}+k_G\left(h_{X,\Omega}\right)^m\right)+ \\ &\quad + \sqrt{\left(\left(I-\mathcal{I}_{X'}\right)\mathcal{S}_X^e R_y, \mathcal{S}_X^e R_y\right)_{w,0}}\,k_G\left(h_{X',\Omega}\right)^m,\end{aligned}$$

for $y \in \overline{\Omega}$ so that

$$\sqrt{\left(\left(I-\mathcal{I}_{X'}\right)\mathcal{S}_X^e R_y, \mathcal{S}_X^e R_y\right)_{w,0}} \leq K'_{\Omega,\kappa}\sqrt{\rho N}+k_G\left(h_{X,\Omega}\right)^m+k_G\left(h_{X',\Omega}\right)^m,$$

and as a consequence of equation 5.37, inequality 5.36 implies

$$\begin{aligned}\left|s_e(y)-s_a(y)\right| &\leq k_G\left(h_{X',\Omega}\right)^m\sqrt{\left(s_e-s_a, s_e\right)_{w,0}} \\ &= k_G\left(h_{X',\Omega}\right)^m\sqrt{\left(\left(I-\mathcal{I}_{X'}\right)\mathcal{S}_X^e R_y, \mathcal{S}_X^e R_y\right)_{w,0}} \\ &\leq k_G\left(h_{X',\Omega}\right)^m\left(K'_{\Omega,\kappa}\sqrt{\rho N}+k_G\left(h_{X,\Omega}\right)^m+k_G\left(h_{X',\Omega}\right)^m\right),\end{aligned}$$

as claimed. ■

Remark 203 *This is the same order of convergence in $h_{X',\Omega}$ that was obtained in Theorem 200 for an arbitrary data function, namely $m = \lfloor\kappa\rfloor$.*

5.7 Type 1 Exact smoother error estimates ($\kappa \geq 0$)

In order to estimate the pointwise error of the Approximate smoother we will need the following Type 1 error estimates of the Exact smoother.

Theorem 204 *(Copy of Theorem 155) Suppose K is a bounded, closed, infinite set and there exist constants C_G, s, $h_G > 0$ such that*

$$G(0)-G(x) \leq C_G\left|x\right|^{2s}, \quad |x| \leq h_G,\ x \in K.$$

Set $k_G = (2\pi)^{-\frac{1}{4}}\sqrt{2C_G}$. Then the Exact smoother $\mathcal{S}_X^e f_d$ satisfies the error estimates

$$\left|f_d(x)-\left(\mathcal{S}_X^e f_d\right)(x)\right| \leq \min\begin{cases}\left\|f_d\right\|_{w,0}\left(\sqrt{\rho N}+k_G\left(h_{X,K}\right)^s\right), \\ \left\|f_d\right\|_{w,0}\sqrt{R_0(0)}, \\ \left\|f_d\right\|_{\infty,K}+\left\|f_d\right\|_{w,0}\sqrt{R_0(0)}\min\left\{1, \frac{R_0(0)}{\rho}\right\},\end{cases} \quad (5.41)$$

and $\mathcal{S}_X^e R_y$ satisfies the double order convergence estimate

$$\left|R_y(x)-\left(\mathcal{S}_X^e R_y\right)(x)\right| \leq \min\begin{cases}\left(\sqrt{\rho N}+k_G\left(h_{X,K}\right)^s\right)^2, \\ R_0(0), \\ \left\|R_y\right\|_{\infty,K}+R_0(0)\min\left\{1, \frac{R_0(0)}{\rho}\right\},\end{cases} \quad (5.42)$$

for $x, y \in K$, when $h_{X,K} = \max\limits_{s \in K}\operatorname{dist}(s, X) \leq h_G$.

Here $R_0(0) = (2\pi)^{-\frac{1}{2}}G(0)$ and X is an independent data set contained in K.

5.8 Type 2 Exact smoother error estimates ($\kappa \geq 1$)

These will just be the Type 1 estimates of the previous section but with the constants from the most appropriate result given in Subsection 4.6.3.

5.9 Convergence of the Approximate smoother to the data function

In Section 5.6 several estimates were derived for the difference between the Approximate smoother s_a and the Exact smoother s_e and a copy of the estimates for the Exact smoother error obtained in Chapter 4 was presented. In this subsection we will combine these estimates by means of the simple triangle inequality: $|f_d(x) - s_a(x)| \leq |f_d(x) - s_e(x)| + |s_e(x) - s_a(x)|$, where f_d is a data function. The convergence results of Subsection 5.6.1 do not involve orders of convergence so they will not be used.

5.9.1 Order-less convergence estimates

The following result is a simple consequence of Corollary 187:

Corollary 205 *Suppose:*

1. *Ω is a bounded, open, connected set and $f_d \in X_w^0$ is a data function.*
2. *$s_e^{(k)}$ is a sequence of Exact smoothers generated by the data $\left[X_k; \widetilde{\mathcal{E}}_{X_k} f_d\right]$ with $X_k \subset \Omega$.*
3. *$s_e^{(k)}$ converges uniformly pointwise to f_d on $\overline{\Omega}$.*
4. *$s_a^{(k)}$ is a sequence of Approximate smoothers generated by $X'_k \subset \Omega$ and $[X_k; f_d(X_k)]$ with the X'_k satisfying $\max\limits_{x \in \Omega} \operatorname{dist}(x, X'_k) \to 0$.*

Then $s_a^{(k)}$ converges uniformly pointwise to f_d on $\overline{\Omega}$.

5.9.2 Error estimates without assuming unisolvent data sets

Type 1 estimates ($\kappa \geq 0$)

The next result gives estimates for an arbitrary data function and a Riesz data function R_y.

Theorem 206 *Suppose:*

1. *the weight function w has property W2 for $\kappa = 0$ and that G is the basis function generated by w. Assume that for some $0 < s < 1$ and $C_G, h_G > 0$ the basis function satisfies*
$$G(0) - G(x) \leq C_G |x|^{2s}, \quad |x| < h_G. \tag{5.43}$$
2. *Let s_e be the Exact smoother generated by the data $\left[X; \widetilde{\mathcal{E}}_X f_d\right]$ for some data function $f_d \in X_w^0$. The independent data X is contained in a data region K which is a closed, bounded, infinite set.*
3. *Suppose that s_a is the Approximate smoother generated by the data $\left[X; \widetilde{\mathcal{E}}_X f_d\right]$ and the points $X' \subset K$.*

Then if $k_G = (2\pi)^{-\frac{d}{4}} \sqrt{2C_G}$ we have the estimates

$$|f_d(x) - s_a(x)| \leq \min \begin{cases} \|f_d\|_{w,0} \left(\sqrt{\rho N} + k_G (h_{X,K})^s + k_G (h_{X',K})^s\right), \\ 2\|f_d\|_{w,0} \sqrt{R_0(0)}, \\ \|f_d\|_{\infty,K} + \|f_d\|_{w,0} (k_G)^2 \frac{(h_{X',K})^{2s}}{\sqrt{\rho}} + \\ \qquad + \|f_d\|_{w,0} \sqrt{R_0(0)} \min\left\{1, \frac{R_0(0)}{\rho}\right\}. \end{cases} \tag{5.44}$$

and

$$\left|R_y(x) - \mathcal{S}^a_{X,X'} R_y(x)\right| \leq \eta_e(\rho) + \min \begin{cases} k_G (h_{X',K})^s \left(\sqrt{\rho N} + k_G (h_{X,K})^s + k_G (h_{X',K})^s\right), \\ R_0(0) \min\left\{1, \frac{R_0(0)}{\rho}\right\}, \\ (k_G)^2 \frac{(h_{X',K})^{2s}}{\sqrt{\rho}} \sqrt{\eta_e(\rho)}. \end{cases} \tag{5.45}$$

where

$$\eta_e(\rho) = \min \begin{cases} \left(\sqrt{\rho N} + k_G (h_{X,K})^s\right)^2, \\ R_0(0), \\ \|R_y\|_{\infty,K} + R_0(0) \min\left\{1, \frac{R_0(0)}{\rho}\right\}. \end{cases} \tag{5.46}$$

when $h_{X,K} = \max\limits_{x\in K} \operatorname{dist}(x, X) \le h_G$ *and* $h_{X',K} = \max\limits_{x\in K} \operatorname{dist}(x, X') \le h_G$.

Proof. From Theorem 204 and Corollary 195,

$$|f_d(x) - s_e(x)| \le \min \begin{cases} \|f_d\|_{w,0} \left(\sqrt{\rho N} + k_G (h_{X,K})^s\right), \\ \|f_d\|_{w,0} \sqrt{R_0(0)}, \\ \|f_d\|_{\infty,K} + \|f_d\|_{w,0} \sqrt{R_0(0)} \min\left\{1, \frac{R_0(0)}{\rho}\right\}. \end{cases}$$

and

$$|s_e(x) - s_a(x)| \le \min \begin{cases} \|f_d\|_{w,0}\, k_G (h_{X',K})^s, \\ \|f_d\|_{w,0} \sqrt{R_0(0)}, \\ \|f_d\|_{w,0} (k_G)^2 \frac{(h_{X',K})^{2s}}{\sqrt{\rho}}. \end{cases}$$

so that

$$\begin{aligned} |f_d(x) - s_a(x)| &\le |f_d(x) - s_e(x)| + |s_e(x) - s_a(x)| \\ &\le \min \begin{cases} \|f_d\|_{w,0} \left(\sqrt{\rho N} + k_G (h_{X,K})^s + k_G (h_{X',K})^s\right), \\ 2\|f_d\|_{w,0} \sqrt{R_0(0)}, \\ \|f_d\|_{\infty,K} + \|f_d\|_{w,0} (k_G)^2 \frac{(h_{X',K})^{2s}}{\sqrt{\rho}} + \\ \qquad + \|f_d\|_{w,0} \sqrt{R_0(0)} \min\left\{1, \frac{R_0(0)}{\rho}\right\}. \end{cases} \end{aligned}$$

Using $\eta(\rho)$ the estimate 5.29 for $\left|\mathcal{S}^e_X R_y(x) - \mathcal{S}^a_{X,X'} R_y(x)\right|$ can be written

$$\begin{aligned} \left|\mathcal{S}^e_X R_y(x) - \mathcal{S}^a_{X,X'} R_y(x)\right| &\le \min \begin{cases} k_G (h_{X',K})^s \left(\sqrt{\rho N} + k_G (h_{X,K})^s + k_G (h_{X',K})^s\right), \\ R_0(0) \min\left\{1, \frac{R_0(0)}{\rho}\right\}, \\ (k_G)^2 \frac{(h_{X',K})^{2s}}{\sqrt{\rho}} \left(\sqrt{\rho N} + k_G (h_{X,K})^s\right), \\ (k_G)^2 \frac{(h_{X',K})^{2s}}{\sqrt{\rho}} \sqrt{R_0(0)}, \\ (k_G)^2 \frac{(h_{X',K})^{2s}}{\sqrt{\rho}} \sqrt{\|R_y\|_{\infty,K} + R_0(0) \min\left\{1, \frac{R_0(0)}{\rho}\right\}}. \end{cases} \\ &\le \min \begin{cases} k_G (h_{X',K})^s \left(\sqrt{\rho N} + k_G (h_{X,K})^s + k_G (h_{X',K})^s\right), \\ R_0(0) \min\left\{1, \frac{R_0(0)}{\rho}\right\}, \\ (k_G)^2 \frac{(h_{X',K})^{2s}}{\sqrt{\rho}} \sqrt{\eta(\rho)}. \end{cases} \end{aligned}$$

Now

$$\left|R_y(x) - \mathcal{S}^a_{X,X'} R_y(x)\right| \le \left|R_y(x) - (\mathcal{S}^e_X R_y)(x)\right| + \left|\mathcal{S}^e_X R_y(x) - \mathcal{S}^a_{X,X'} R_y(x)\right|,$$

and since the estimate for $|R_y(x) - (\mathcal{S}^e_X R_y)(x)|$ from 5.42 can be written $|R_y(x) - (\mathcal{S}^e_X R_y)(x)| \le \eta(\rho)$ we have 5.45. ∎

Remark 207 *Define* $h_a = \left((h_{X,K})^s + (h_{X',K})^s\right)^{1/s}$. *Then:*

1. *From 5.44*

$$|f_d(x) - s_a(x)| \le \|f_d\|_{w,0} \left(\sqrt{\rho N} + k_G (h_a)^s\right), \quad x \in K,$$

and the order of convergence is s *in* h_a.

2. *From 5.45: for $x, y \in K$*

$$\begin{aligned}\left|R_y(x) - \mathcal{S}^a_{X,X'} R_y(x)\right| &\leq \left(\sqrt{\rho N} + k_G\left(h_{X,K}\right)^s\right)^2 + \sqrt{\rho N} k_G\left(h_{X',K}\right)^s + \\ &\quad + \left(k_G\right)^2\left(h_{X',K}\right)^s\left(h_{X',K}\right)^s + \left(k_G\right)^2\left(h_{X,K}\right)^{2s} \\ &\leq \left(\sqrt{\rho N} + k_G\left(h_{X,K}\right)^s + k_G\left(h_{X',K}\right)^s\right)^2 \\ &= \left(\sqrt{\rho N} + k_G\left(h_a\right)^s\right)^2,\end{aligned}$$

and the order of convergence is $2s$ in h_a.

3. *The last two Approximate smoother error estimates are analogous to the Exact smoother error estimates*

$$\left|f_d(x) - \left(\mathcal{S}^e_X f_d\right)(x)\right| \leq \|f_d\|_{w,0}\left(\sqrt{\rho N} + k_G\left(h_{X,K}\right)^s\right),$$

and

$$\left|R_y(x) - \left(\mathcal{S}^e_X R_y\right)(x)\right| \leq \left(\sqrt{\rho N} + k_G\left(h_{X,K}\right)^s\right)^2,$$

from 5.41 and 5.42.

Type 2 error estimates ($\kappa \geq 1$)

We want to derive Type 2 error estimates for $|f_d(x) - s_a(x)|$. In fact, these results can be proved by combining the Type 2 Exact smoother error estimates from Subsection 4.6.3 with the Type 2 estimates for $|s_e(x) - s_a(x)|$ from Subsubsection 5.6.3.

Theorem 208 *Suppose a weight function satisfies property W2 for some $\kappa \geq 1$ and denote the basis function by G. Then the Approximate smoother error estimates 5.28 and 5.29 hold for*

$$C_G = -\frac{1}{2}\left(\nabla^2 G\right)(0)\, d, \quad s = 1, \quad h_G = \infty, \quad k_G = (2\pi)^{-\frac{d}{4}}\sqrt{-\left(\nabla^2 G\right)(0)}\sqrt{d}.$$

However, if the weight function is radial we can use the estimates of:

Theorem 209 *Suppose a radial weight function satisfies property W2 for $\kappa = 1$ and denote the (radial) basis function by G. Set $r = |x|$. Then:*

1. *If $G(x) = f\left(r^2\right)$ the Approximate smoother error estimates of Theorem 206 hold for*

$$C_G = -f'(0)\, d^2, \quad s = 1, \quad h_G = \infty, \quad k_G = (2\pi)^{-\frac{d}{4}}\sqrt{-2f'(0)}d.$$

2. *If $G(x) = g(r)$ then the Approximate smoother error estimates of Theorem 206 hold for*

$$C_G = -\frac{1}{2}g''(0)\, d^2, \quad s = 1, \quad h_G = \infty, \quad k_G = (2\pi)^{-\frac{d}{4}}\sqrt{-g''(0)}d.$$

If the weight function is a tensor product we can use the estimates of:

Theorem 210 *Suppose a tensor product weight function satisfies property W2 for $\kappa = 1$ and denote the univariate basis function by G_1. Then the Approximate smoother error estimates of Theorem 206 hold when*

$$C_G = -\frac{d}{2}G_1(0)^{d-1} D^2 G_1(0), \quad s = 1, \quad h_G = \infty, \quad k_G = (2\pi)^{-\frac{d}{4}}\sqrt{-G_1(0)^{d-1} D^2 G_1(0)}\sqrt{d}.$$

5.9.3 Estimates using unisolvent data sets

The next result is an error estimate for an arbitrary data function in X_w^0.

Theorem 211 *Suppose:*

1. *w is a weight function with property W2 for some $\kappa \geq 1$ and let G be the basis function generated by w. Set $m = \lfloor \kappa \rfloor$.*
2. *s_e is the Exact smoother generated by the data $\left[X; \widetilde{\mathcal{E}}_X f_d\right]$ for some data function $f_d \in X_w^0$, and X is contained in the data region Ω.*
3. *s_a is the Approximate smoother generated by $\left[X; \widetilde{\mathcal{E}}_X f_d\right]$ and the set $X' \subset \Omega$.*
4. *We use the notation and assumptions of Lemma 72 which means here that X' and X are m-unisolvent and that Ω is a bounded, open, connected set whose boundary satisfies the cone condition.*

Then there exist positive $h_{\Omega,m}, c_{\Omega,m}, K'_{\Omega,m}$ and $k_G = \frac{d^{m/2}}{(2\pi)^{d/2}} \left(c_{\Omega,m}\right)^m K'_{\Omega,m} \max\limits_{|\beta|=m} \left|D^{2\beta} G(0)\right|$ such that

$$\left|f_d(x) - s_a(x)\right| \leq \sqrt{\left(f_d - s_e, f_d\right)_{w,0}} \left(K'_{\Omega,m}\sqrt{\rho N} + k_G \left(h_{X,\Omega}\right)^m\right) + \\ + \sqrt{\left(s_e - s_a, s_e\right)_{w,0}} k_G \left(h_{X',\Omega}\right)^m, \quad x \in \overline{\Omega},$$

when $h_{X,K} = \sup\limits_{x\in\Omega} \operatorname{dist}(x, X) \leq h_G$ and $h_{X',K} = \sup\limits_{x\in\Omega} \operatorname{dist}(x, X') \leq h_G$.

Further, $\sqrt{\left(s_e - s_a, s_e\right)_{w,0}} \leq \left\|s_e\right\|_{w,0} \leq \left\|f_d\right\|_{w,0}$ and $\sqrt{\left(f_d - s_e, f_d\right)_{w,0}} \leq \left\|f_d\right\|_{w,0}$.

Proof. From Theorem 200

$$\left|s_e(x) - s_a(x)\right| \leq \sqrt{\left(s_e - s_a, s_e\right)_{w,0}} k_G \left(h_{X',\Omega}\right)^m, \quad x \in \overline{\Omega},$$

when $h_{X',\Omega} < h_{\Omega,m}$. From inequality 4.31

$$\left|f_d(x) - s_e(x)\right| < \sqrt{\left(f_d - s_e, f_d\right)_{w,0}} \left(K'_{\Omega,m}\sqrt{\rho N} + k_G \left(h_{X,\Omega}\right)^m\right), \quad x \in \overline{\Omega},$$

when $h_{X,\Omega} \leq h_{\Omega,m}$. Hence when $x \in \overline{\Omega}$

$$\begin{aligned} \left|f_d(x) - s_a(x)\right| &\leq \left|f_d(x) - s_e(x)\right| + \left|s_e(x) - s_a(x)\right| \\ &\leq \sqrt{\left(f_d - s_e, f_d\right)_{w,0}} \left(K'_{\Omega,m}\sqrt{\rho N} + k_G \left(h_{X,\Omega}\right)^m\right) + \\ &\quad + \sqrt{\left(s_e - s_a, s_e\right)_{w,0}} k_G \left(h_{X',\Omega}\right)^m. \end{aligned}$$

The final inequalities follow from Theorem 200 and Corollary 153. ■

Remark 212 *We can combine the two density measures $h_{X,\Omega}$ and $h_{X',\Omega}$ into one measure*

$$h_a = \left(\left(h_{X,\Omega}\right)^m + \left(h_{X',\Omega}\right)^m\right)^{1/m}.$$

From Theorem 206 we have $\sqrt{\left(f_d - s_e, f_d\right)_{w,0}} \leq \left\|f_d\right\|_{w,0}$ and $\sqrt{\left(s_e - s_a, s_e\right)_{w,0}} \leq \left\|f_d\right\|_{w,0}$ so that

$$\begin{aligned} \left|f_d(x) - s_a(x)\right| &\leq \sqrt{\left(f_d - s_e, f_d\right)_{w,0}} \left(K'_{\Omega,m}\sqrt{\rho N} + k_G \left(h_{X,\Omega}\right)^m\right) + \\ &\quad + \sqrt{\left(s_e - s_a, s_e\right)_{w,0}} k_G \left(h_{X',\Omega}\right)^m \\ &\leq \left\|f_d\right\|_{w,0} \left(K'_{\Omega,m}\sqrt{\rho N} + k_G \left(h_{X,\Omega}\right)^m + k_G \left(h_{X',\Omega}\right)^m\right) \\ &\leq \left\|f_d\right\|_{w,0} \left(K'_{\Omega,m}\sqrt{\rho N} + k_G \left(h_a\right)^m\right), \end{aligned}$$

so that when $\rho = 0$ the order of convergence in h_a is $m = \lfloor \kappa \rfloor$.

An important application of the Approximate smoother is the case where X' is a regular rectangular grid. To apply the last theorem we must show that if the grid size decreases then X' eventually contains a m-unisolvent subset. The next theorem confirms this. The result of the next theorem is obvious in one-dimension but results for polynomials in higher dimensions can be complex and/or unexpected. However, here things are OK. We will need the following lemma:

Lemma 213 *We have the following unisolvency results:*

1. *The set* $\left\{\gamma \in \mathbb{Z}^d : 0 \leq \gamma < n\right\}$ *is unisolvent w.r.t.* P_n.
2. *Translations of minimal unisolvent sets are minimal unisolvent sets.*
3. *Dilations of minimal unisolvent sets are minimal unisolvent sets.*

Proof. Part 1. From the definition of unisolvency, Definition 66, we must show that for each $p \in P_n$, $p(\gamma) = 0$ for $0 \leq \gamma < n$ implies $p = 0$. The proof will be by induction on the order of the polynomial.

Clearly the lemma is true for $n = 1$ since P_1 is the constant polynomials.

Now assume that $n \geq 2$ and that if $p \in P_n$ and $p(\gamma) = 0$ for $0 \leq \gamma < n$ then $p = 0$. Set $p(x) = \sum\limits_{|\beta|<n} c_\beta x^\beta$. Then if $\gamma_k = 0$ and $\gamma_i = 1$ when $i \neq k$ then $0 = \sum\limits_{|\beta|<n} c_\beta \gamma^\beta$ implies $c_\beta = 0$ when $\beta_k = 0$. Thus $p(\gamma) = 0$ for $0 \leq \gamma < n$ implies $c_\beta = 0$ when $\beta_i = 0$ for some i. Consequently, $p = 0$ if $n \leq d$ else $p(x) = \sum\limits_{\substack{|\beta|<n \\ \beta>0}} c_\beta x^\beta$. If $n > d$ we can write $p(x) = x^{\mathbf{1}} q(x)$ where $q \in P_{n-1}$, so that $q(\gamma) = 0$ for $1 \leq \gamma < n$ must imply $q = 0$. Finally, if we define $r \in P_{n-1}$ by $r(x) = q(x+1)$ then $r(\gamma) = 0$ for $0 \leq \gamma < n-1$ must imply $r = 0$, and so the truth of our lemma for $n-1$ implies the truth of the lemma for n, and the lemma is proved.

Parts 2 and 3. From the definition of a cardinal basis, Definition 68, there is a unique cardinal basis $\{l_i\}$ of P_n associated with a set $A = \{a_i\}$ iff A is minimally unisolvent of order n. Now by definition $l_i(a_j) = \delta_{i,j}$. Hence, if $\tau, \delta \in \mathbb{R}^d$ and δ has positive components, then the cardinal basis associated with the translation $A + \tau$ is $\{l_i(\cdot - \tau)\}$ and the cardinal basis associated with the dilation δA is $\{l_i(\cdot/\delta)\}$. ■

Theorem 214 *Suppose* $X' = \left\{x'_\alpha = a + h\alpha \mid \alpha \in \mathbb{Z}^d \text{ and } 0 \leq \alpha < \mathcal{N}'\right\}$ *is the regular, rectangular grid introduced in Definition 170.*

Then X' is m-unisolvent if $\mathcal{N}' \geq m$.

Proof. Since $(X' - a)/h = \left\{\alpha \mid \alpha \in \mathbb{Z}^d \text{ and } 0 \leq \alpha < \mathcal{N}'\right\}$ and $\mathcal{N}' \geq m$, part 1 of the lemma implies $(X' - a)/h$ is m-unisolvent and thus from the definition of unisolvency, Definition 66, $(X' - a)/h$ must contain a minimal unisolvent subset. Parts 2 and 3 of the lemma imply that X' contains a minimal unisolvent subset and so X' is unisolvent. ■

We can improve the last convergence result when the data functions have the form R_y.

Theorem 215 *Suppose:*

1. *w is a weight function with property W2 for some $\kappa \geq 1$ and let G be the basis function generated by w. Set $m = \lfloor \kappa \rfloor$.*
2. *$\mathcal{S}^e_X R_y$ is the Exact smoother generated by the data $\left[X; \widetilde{\mathcal{E}}_X R_y\right]$ for some data function R_y, and X is contained in the data region Ω.*
3. *$\mathcal{S}^a_{X,X'} R_y$ is the Approximate smoother generated by $\left[X; \widetilde{\mathcal{E}}_X R_y\right]$ and the set X' is contained in the data region Ω.*
4. *We use the notation and assumptions of Lemma 72 which means here that X' and X are m-unisolvent and that Ω is a bounded, open, connected set whose boundary satisfies the cone condition.*

Then there exist positive $h_{\Omega,m}, c_{\Omega,m}, K'_{\Omega,m}$ and $k_G = \frac{d^{m/2}}{(2\pi)^{d/2}} (c_{\Omega,m})^m K'_{\Omega,m} \max\limits_{|\beta|=m} \left|D^{2\beta} G(0)\right|$ such that

$$\left|R_y(x) - \mathcal{S}^a_{X,X'} R_y(x)\right| \leq \left(K'_{\Omega,m} \sqrt{\rho N} + k_G (h_{X,\Omega})^m + k_G (h_{X',\Omega})^m\right)^2,$$

for $x, y \in \overline{\Omega}$, when $h_{X,\Omega} = \sup\limits_{x \in \Omega} \operatorname{dist}(x, X) \leq h_G$ and $h_{X',\Omega} = \sup\limits_{x \in \Omega} \operatorname{dist}(x, X') \leq h_G$.

Proof. When $h_{X,\Omega} \le h_{\Omega,m}$ and $h_{X',\Omega} \le h_{\Omega,m}$ Theorem 202 implies

$$\left|\mathcal{S}^e_X R_y(x) - \mathcal{S}^a_{X,X'} R_y(x)\right| \le k'_G (h_{X',\Omega})^m \left(K'_{\Omega,m}\sqrt{\rho N} + k_G (h_{X,\Omega})^m + k_G (h_{X',\Omega})^m\right),$$

for $x \in \overline{\Omega}$, and 5.42 implies that

$$|R_y(x) - \mathcal{S}^e_X R_y(x)| \le \left(K'_{\Omega,m}\sqrt{\rho N} + k_G (h_{X,\Omega})^m\right)^2, \quad x, y \in \overline{\Omega}.$$

Hence when $x \in \overline{\Omega}$

$$\begin{aligned}
\left|R_y(x) - \mathcal{S}^a_{X,X'} R_y(x)\right| &\le |R_y(x) - \mathcal{S}^e_X R_y(x)| + \left|\mathcal{S}^e_X R_y(x) - \mathcal{S}^a_{X,X'} R_y(x)\right| \\
&\le \left(K'_{\Omega,m}\sqrt{\rho N} + k_G (h_{X,\Omega})^m\right)^2 + \\
&\qquad k'_G (h_{X',\Omega})^m \left(K'_{\Omega,m}\sqrt{\rho N} + k_G (h_{X,\Omega})^m + k_G (h_{X',\Omega})^m\right) \\
&< \left(K'_{\Omega,m}\sqrt{\rho N} + k_G (h_{X,\Omega})^m + k_G (h_{X',\Omega})^m\right)^2.
\end{aligned}$$

■

Remark 216 *We have*

$$\left|R_y(x) - \mathcal{S}^a_{X,X'} R_y(x)\right| \le \left(K'_{\Omega,m}\sqrt{\rho N} + k_G (h_a)^m\right)^2, \quad x, y \in \overline{\Omega},$$

where $h_a = \left((h_{X,\Omega})^m + (h_{X',\Omega})^m\right)^{1/m}$ *so that when* $\rho = 0$ *the order of convergence is* $2m = 2\lfloor\kappa\rfloor$. *This compares with order* $m = \lfloor\kappa\rfloor$ *obtained in Remark 212 in the case of an arbitrary data function.*

5.10 Numerical experiments with scaled, extended B-splines (non-unisolvency)

In this section the convergence of the Approximate smoother to its data function is studied numerically using scaled, extended B-splines, the first of which will be the hat function. We will use the same 1-dimensional data functions and B-spline basis functions 4.37 that were used for the numerical interpolation experiments described in Section 4.8. We will only consider the numerical experiments in one dimension to allow the data density parameter $h_{X,K} = \max\limits_{x \in K} \operatorname{dist}(x, X)$ to be easily calculated.

It is easier to validate the non-unisolvent error estimates i.e. those derived without assuming unisolvent data sets. This is because in the non-unisolvent case the constants can be known precisely. In the case of unisolvent data the theory is complex and it is unclear what are suitable upper bounds for the constants $K'_{\Omega,m}$, $c_{\Omega,m}$, $h_{\Omega,m}$ in Theorem 211.

Because all the extended B-spline basis weight functions have a power of $\sin x$ in the denominator we will use the special classes of data functions developed in Section 2.6. The multivariate theory of local data spaces $X^0_w(\Omega)$ where Ω is a bounded open set was developed in Sections 2.7 and 3.5.

The 1-dimensional data set is constructed using a uniform distribution on the data region $K = [-1.5, 1.5]$.

We now assume that X' is a regular grid so that

$$h_{X',K} = \frac{3}{2N'}, \tag{5.47}$$

where $N' \ge 2$ is the number of grid cells.

Each of 20 data files is exponentially sampled using a multiplier of approximately 1.3 and then we plot $\log_{10} h_{X,K}$ against $\log_{10} N$ where $N = |X|$. It then seems quite reasonable to use a least-squares linear fit and in this case we obtain

$$h_{X,K} \simeq 3.09 N^{-0.81}. \tag{5.48}$$

For ease of calculation let

$$h_{X,K} = h_1 N^{-a}, \quad h_1 = 3.09, \quad a = 0.81. \tag{5.49}$$

Noting the error estimates of Theorem 206 we use 5.49 to write $h_{X,K}^s = h_1^s N^{-as}$ and define the Exact smoother error estimates

$$|f_d(x) - (\mathcal{S}_X^e f_d)(x)| \lesssim E_N(\rho), \quad x \in K, \tag{5.50}$$

and

$$|R_y(x) - (\mathcal{S}_X^e R_y)(x)| \lesssim \varepsilon_N(\rho), \quad x, y \in K, \tag{5.51}$$

by

$$E_N(\rho) = \min \begin{cases} \|f_d\|_{w,0} \left(\sqrt{\rho N} + k_G \frac{h_1^s}{N^{as}} + k_G \left(\frac{3}{2N'}\right)^s\right), \\ \|f_d\|_{w,0}\, 2\sqrt{R_0(0)}, \\ \|f_d\|_{\infty,K} + \|f_d\|_{w,0} (k_G)^2 \left(\frac{3}{2N'}\right)^{2s} \frac{1}{\sqrt{\rho}} + \\ \qquad + \|f_d\|_{w,0} \sqrt{R_0(0)} \min\left\{1, \frac{R_0(0)}{\rho}\right\}. \end{cases} \tag{5.52}$$

and

$$\varepsilon_N(\rho) = \eta_e(\rho) + \min \begin{cases} k_G \left(\frac{3}{2N'}\right)^s \left(\sqrt{\rho N} + k_G \frac{h_1^s}{N^{as}} + k_G \left(\frac{3}{2N'}\right)^s\right), \\ R_0(0) \min\left\{1, \frac{R_0(0)}{\rho}\right\}, \\ (k_G)^2 \left(\frac{3}{2N'}\right)^{2s} \sqrt{\frac{\eta_e(\rho)}{\rho}}. \end{cases} \tag{5.53}$$

where

$$\eta_e(\rho) = \min \begin{cases} \left(\sqrt{\rho N} + k_G \frac{h_1^s}{N^{as}}\right)^2, \\ R_0(0), \\ \|R_y\|_{\infty,K} + R_0(0) \min\left\{1, \frac{R_0(0)}{\rho}\right\}. \end{cases} \tag{5.54}$$

5.10.1 Extended B-splines with $n = 1$

<u>The case $n = 1$, $l = 1$</u>

The 1-dimensional hat weight function w_Λ is given by $w_\Lambda(\xi) = \frac{\xi^2}{\sin^2 \xi}$ and was discussed in Subsections 1.2.1 and 1.2.6. This is a scaled, extended B-spline basis function corresponding to $n = l = 1$ and thus Theorem 7 and part 6 Remark 2 imply $\kappa < 1/2$.

Now suppose Π is the 1-dimensional unit-valued rectangular function with support $[-.5, .5]$. Then it was shown in Subsubsection 2.6.1 that:

Theorem 217 *If $u \in L^2(\mathbb{R}^1)$ then $u * \Pi$ is a data function i.e. $u * \Pi \in X_{w_\Lambda}^0$ where w_Λ is the hat weight function. Further, if we define $V \in L_{loc}^1 \cap C_{BP}^{(0)}$ by*

$$V(x) = \int_0^x u(t)\, dt, \tag{5.55}$$

then $DV = u$ a.e.,

$$u * \Pi = (2\pi)^{-\frac{1}{2}} \left(V\left(x + \frac{1}{2}\right) - V\left(x - \frac{1}{2}\right)\right),$$

and

$$\|u * \Pi\|_{w_\Lambda,0} = (2\pi)^{-\frac{d}{4}} \|u\|_2 .$$

To obtain our data function $f_d = u * \Pi$ we will choose

$$u(x) = e^{-x^2},$$

for which

$$V = (2\pi)^{\frac{1}{2}} \operatorname{erf}, \qquad \|u\|_2 = 2(2\pi)^{\frac{1}{4}}, \qquad \|u * \Pi\|_{w_\Lambda,0} = 2, \tag{5.56}$$

and

$$f_d(x) = u * \Pi(x) = \operatorname{erf}\left(x + \frac{1}{2}\right) - \operatorname{erf}\left(x - \frac{1}{2}\right). \tag{5.57}$$

Since we are using the hat basis function, from Subsubsection 2.6.1

$$G(0) = 1, \quad C_G = 1, \quad s = 1/2, \quad h_G = \infty, \quad k_G = (2\pi)^{-1/4}\sqrt{2},$$

and from 5.56 and 5.57,

$$\|f_d\|_{w,0} = 2.$$

Finally, $Df_d(x) = (2\pi)^{\frac{1}{2}} e^{-\left(x+\frac{1}{2}\right)^2} - (2\pi)^{\frac{1}{2}} e^{-\left(x-\frac{1}{2}\right)^2} = 0$ iff $x = 0$ so the maximum value of the data function on the data region $K = [-1.5, 1.5]$ is

$$\max_K f_d = \operatorname{erf}\left(\frac{1}{2}\right) - \operatorname{erf}\left(-\frac{1}{2}\right) \simeq 1.041.$$

For the *double rate* convergence experiment we will use the Riesz data function $R_0 = (2\pi)^{-\frac{1}{2}}\Lambda$.

Numerical results

Absolute smoother error vs. Smoothing parameter We start by plotting the absolute error bounds given by 5.52, 5.53 and 5.54, together with the actual absolute smoother errors, against the smoothing parameter. To be precise the actual errors are averaged over a regular *error grid* covering the data region. The results for the data function f_d are shown in Figure 5.1 and the results for the Riesz data function R_0 are shown in Figure 5.2:

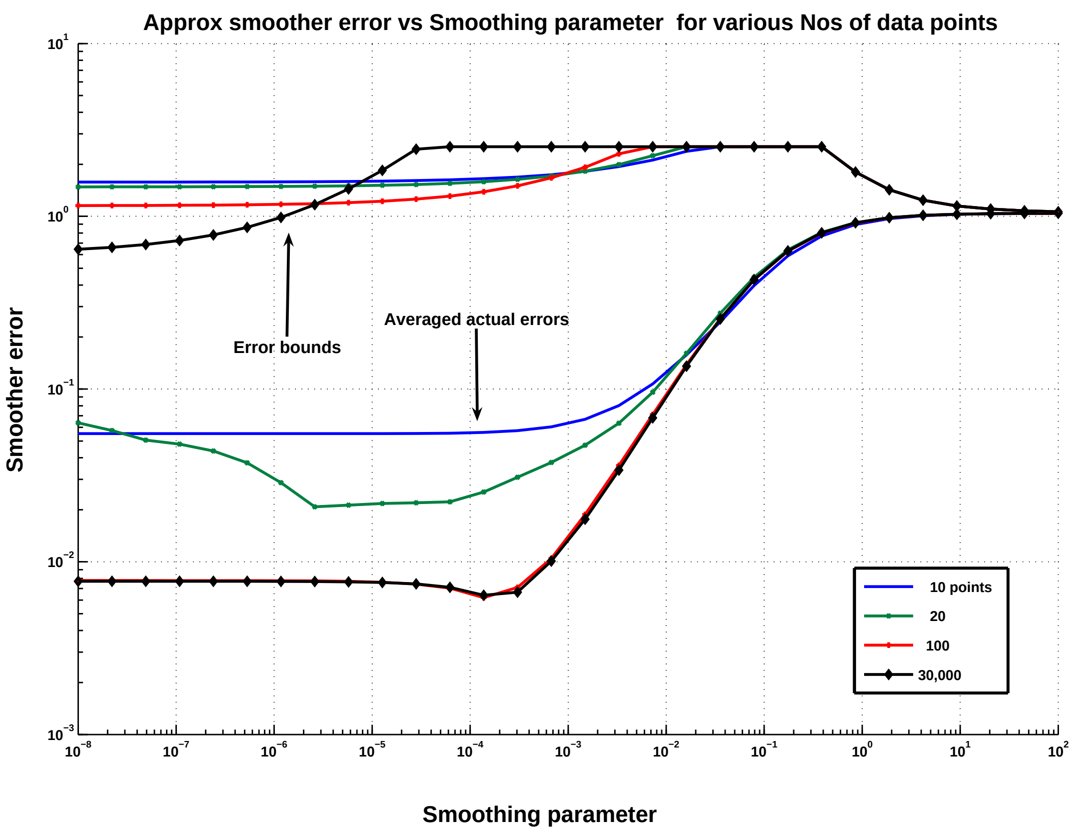

FIGURE 5.1. Data function: $\operatorname{erf}\left(x+\frac{1}{2}\right) - \operatorname{erf}\left(x-\frac{1}{2}\right)$. Num smoother grid cells is 16.

Smoother error on Data region As the number of data points increases the smoother error of f_d is observed to take the form of stable alternating positive/negative spikes as shown in Figure 5.3 which corresponds to 30K data points. The positive errors are modulated by a hat function and the zero error points have a period of 3/16; note there are 16 cells in the smoother grid.

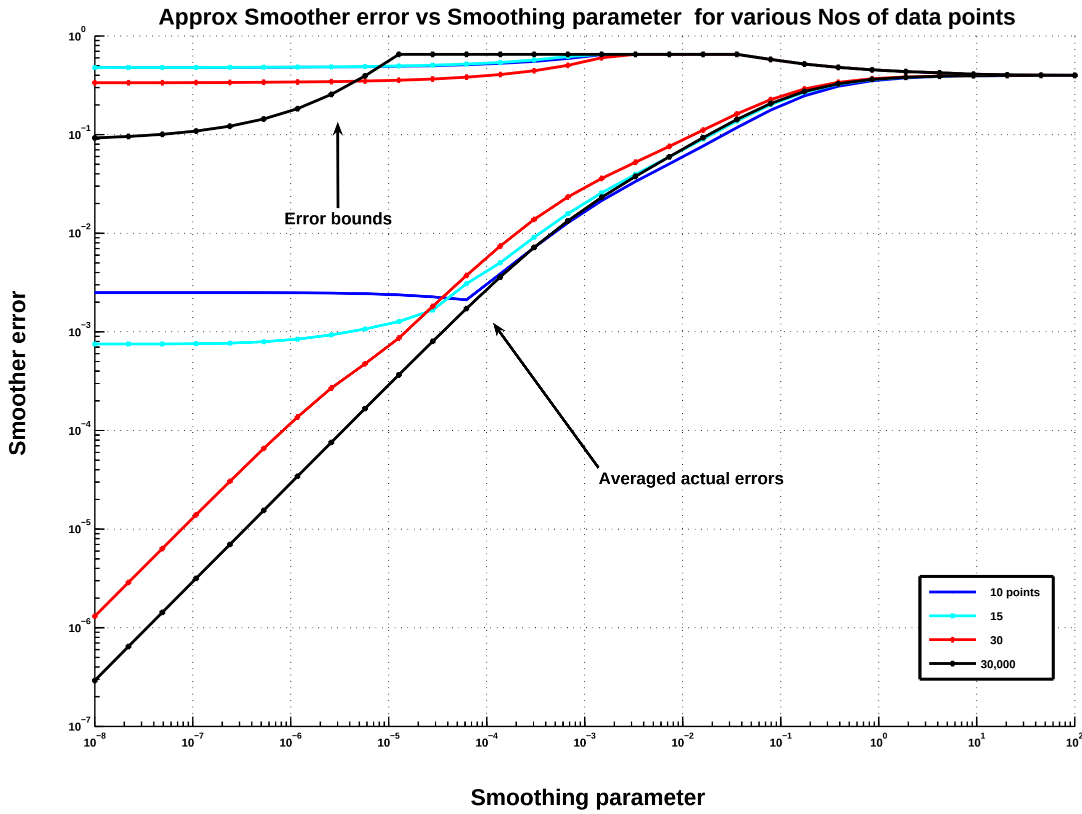

FIGURE 5.2. Riesz data function: $R_0 = \Lambda$ (Hat func). Num smoother grid cells is 16.

It is seen in Figure 5.4 that as the number of points increases the smoother error of the Riesz data function R_0 is observed to stabilize to a wide negative central spike surrounded by smaller spikes of about 1/5th the amplitude.

It is clear that, except for smoothing parameters larger than 10, the theoretical error bounds given in 5.52 and 5.53 capture only a small part of the actual error. The convergence of the error bounds to the average spatial errors for large ρ is typical. The averaging process also disguises the instability indicated by Figure 5.5.

Absolute smoother error vs. Data density Following the numerical work of Chapter 4 which studied the Exact smoother we will also apply the 'spike' filter to the Approximate smoother errors. This filter calculates the value below which 80% of the errors lie. The filter is designed to remove 'large', isolated spikes which dominate the actual smoother errors. The smoother error was calculated on a grid with 200 cells applied to the domain of the data function. No filter is used for the first five interpolants because there is no instability for small numbers of data points. This filter will be used in all the cases below.

We now move on to examine the relationship between the smoother error and the data density. To do this we need to fix a value for the smoothing coefficient and here we choose $\rho = 10^{-6}$. We select this value because often the Exact smoother average error curve has a minimum near this value. Using the functions and parameters discussed at the end of the last subsection we obtain the four subplots displayed in Figure 5.5, each subplot being the superposition of 20 interpolants.

The two upper subplots relate to the data function 5.57 and the lower subplots relate to the data function R_0. The right-hand subplots are filtered versions of the actual errors on the right. The (blue) points above the actual errors in Figure 5.5 are the estimated upper bounds for the error given by the theoretical estimates 5.44 and 5.45. The annotation at the bottom of the figure supplies the following additional information:

Input parameters

$\mathbf{N = L = 1}$ **-** the hat function is a member of the family of scaled, extended splines with the indicated parameter values.

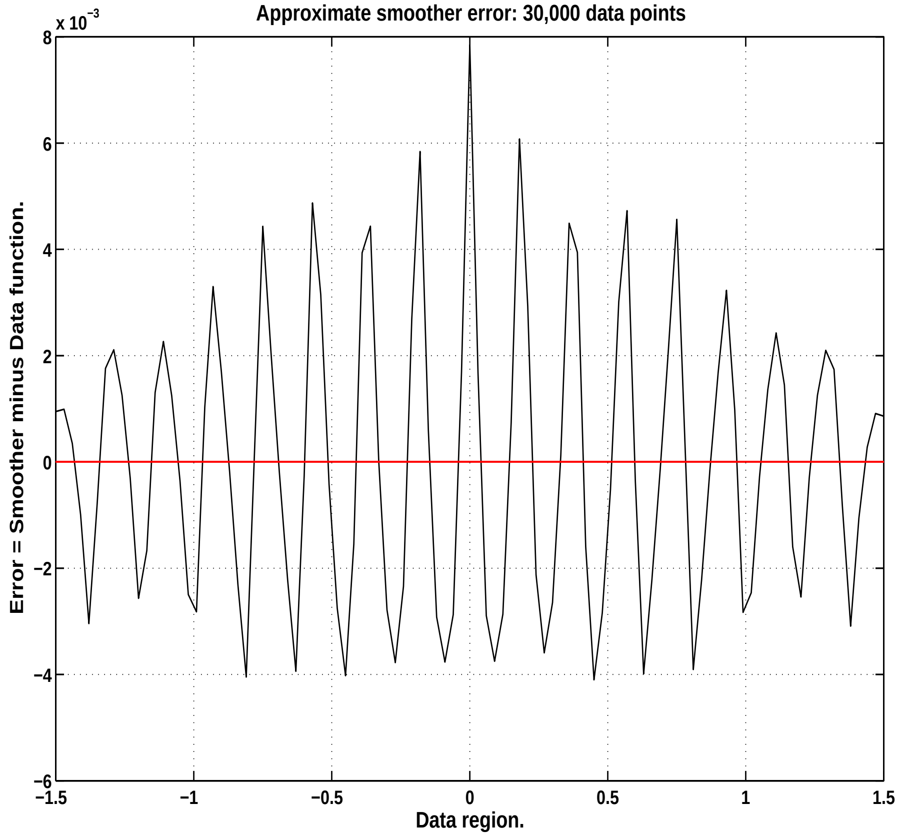

FIGURE 5.3. Data function: $\operatorname{erf}\left(x+\frac{1}{2}\right)-\operatorname{erf}\left(x-\frac{1}{2}\right)$. Number of smoother grid cells is 16.

spl scale 1/2 - the actual scaling of the spline basis function is 1/2 divided by **spl scale**.

sm parm 1e-6 - the smoothing parameter is zero for interpolation.

samp 20 - the sample size which is the number of test data files generated. The data function is evaluated on the data interval $K=[-1.5,1.5]$ at points selected using a uniform (statistical) distribution.

pts 2:50K - specifies the smallest number of data points to be 2 and the largest number of data points to be 30,000. The other values are given in exponential steps with a multiplier of approximately 1.2.

sm grid cell 16 - number of cells in the regular smoother grid superimposed on the data region K.

Output parameters/messages

No ill-condit - this indicates all Exact smoother matrices were always properly conditioned.

Note that all the plots of error vs. data density given below have the same format and annotations.

Clearly both theoretical error bounds substantially underestimates the convergence rate. The instabilities associated with small numbers of data points disappear as the number of data points increases. The smoother of R_0 is less stable than that of f_d.

The case $n=1$, $l=2$

Here we are interested in the 1-dimensional extended B-spline basis function with $n=1$ and $l=2$. The following result, which is Theorem 77 in one dimension, will allow us to generate data functions for which the X_w^0 norm can be calculated This result is closely related to the calculations done above for the hat function.

Theorem 218 ***Data functions*** *Suppose w is the B-spline tensor product weight function with parameters n and l and univariate 4.35, and that $U\in L^2\left(\mathbb{R}^1\right)$ with $D^nU\in L^2\left(\mathbb{R}^1\right)$ in the sense of distributions. Then if we define the distribution*

$$f_d=\delta_2^l U,\quad l=1,2,3,\ldots,$$

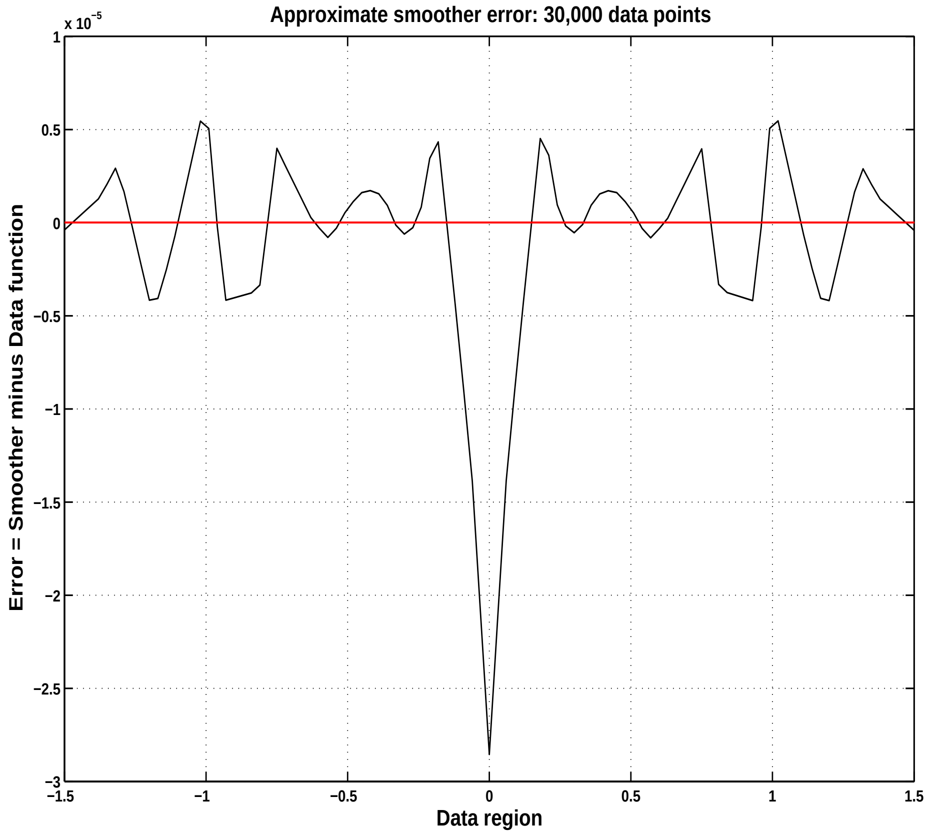

FIGURE 5.4. Riesz data function: $R_0 = \Lambda$ (hat func). Number of smoother grid cells is 16.

where δ_2 is the central difference operator

$$\delta_2 U = U(\cdot + 1) - U(\cdot - 1), \tag{5.58}$$

it follows that $f_d \in X_w^0$ and

$$\|f_d\|_{w,0} = 2^l \|D^n U\|_2 . \tag{5.59}$$

Our tensor product basis function G has univariate 4.37 i.e.

$$\begin{aligned} G(t) &= (-1)^{l-n} \tfrac{(2\pi)^{l/2}}{2^{2(l-n)+1}} \left(D^{2(l-n)} \left((*\Lambda)^l \right) \right) \left(\tfrac{t}{2} \right) \\ &= -\tfrac{\pi}{4} \left(D^2 (\Lambda * \Lambda) \right) \left(\tfrac{t}{2} \right), \quad t \in \mathbb{R}^1. \end{aligned}$$

But

$$\begin{aligned} D^2 (\Lambda * \Lambda) = \Lambda * D^2 \Lambda &= \Lambda * (\delta(\cdot + 1) - 2\delta + \delta(\cdot - 1)) \\ &= \frac{1}{\sqrt{2\pi}} (\Lambda(\cdot + 1) - 2\Lambda + \Lambda(\cdot - 1)), \end{aligned}$$

so that

$$G(t) = -\frac{\sqrt{2\pi}}{8} \left(\Lambda\left(\tfrac{t}{2} + 1\right) - 2\Lambda\left(\tfrac{t}{2}\right) + \Lambda\left(\tfrac{t}{2} - 1\right) \right),$$

and hence

$$DG(t) = -\frac{\sqrt{2\pi}}{16} \left(\Lambda'\left(\tfrac{t}{2} + 1\right) - 2\Lambda'\left(\tfrac{t}{2}\right) + \Lambda'\left(\tfrac{t}{2} - 1\right) \right),$$

i.e. $\|DG\|_\infty = \frac{3}{16}\sqrt{2\pi}$. Since the (distributional) derivative is bounded the distributional Taylor series expansion of Lemma 42 can be used to write

$$G(0) - G(t) \le \|DG\|_\infty |t|, \quad x \in \mathbb{R}^1,$$

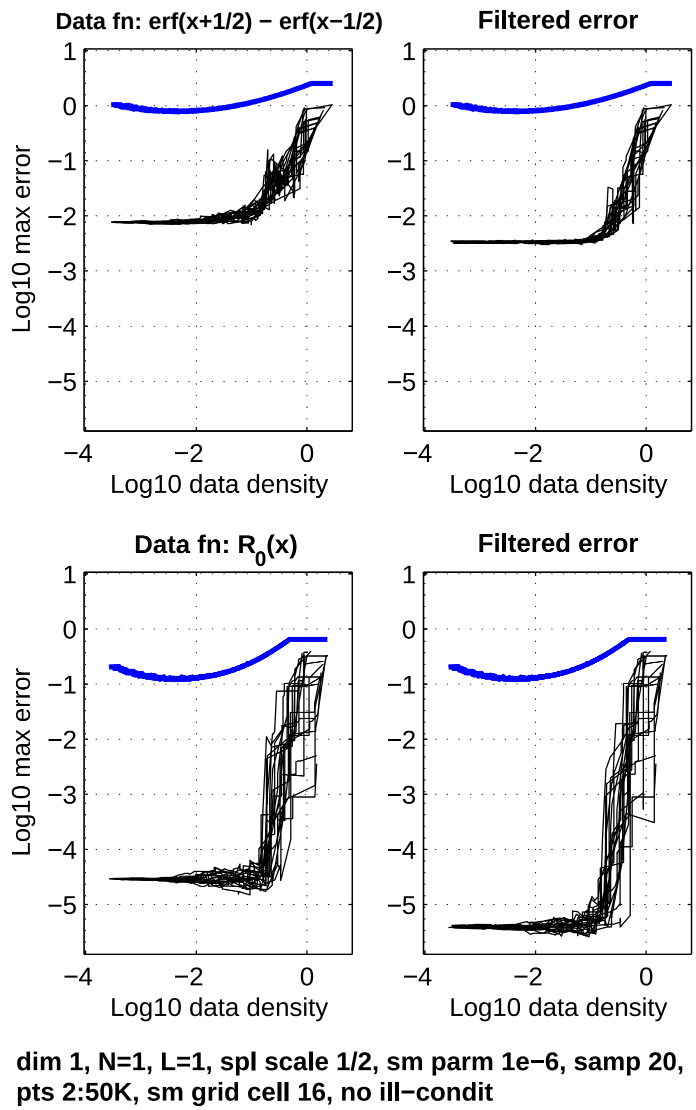

FIGURE 5.5. Error of Approximate smoother vs data density.

which means that

$$G(0) = \frac{\sqrt{2\pi}}{4},\ C_G = \|DG_{1,2}\|_\infty = \frac{3}{16}\sqrt{2\pi},\ s = \frac{1}{2},\ h_G = \infty.$$

With reference to the last theorem we will choose the bell-shaped data function

$$f_d = \delta_2^2 U \in X_w^0, \tag{5.60}$$

where

$$U(x) = \frac{e^{-k_{1,2}x^2}}{\delta_2^2\left(e^{-k_{1,2}x^2}\right)(0)} = \frac{e^{-k_{1,2}x^2}}{2\left(1-e^{-4k_{1,2}}\right)}, \quad k_{1,2} = 0.3,$$

so that $\|f_d\|_{\infty,K} = f_d(0) = 1$ and

$$\|f_d\|_{w,0} = 4\|DU\|_2 = \sqrt[4]{2\pi}\frac{\sqrt[4]{4k_{1,2}}}{1-e^{-4k_{1,2}}}.$$

The error estimates are given by 5.44 and 5.45 where $k_G = (2\pi)^{-\frac{1}{4}}\sqrt{2C_G}$ and $\|R_0\|_{\infty,K} = R_0(0) = (2\pi)^{-\frac{1}{2}}G(0) = 1/4$.

For the theory developed in the previous sections it was convenient to use the unscaled B-spline weight function definition 4.35 but for computations it may be easier to use the unscaled version of the extended B-spline basis function given in the next theorem.

Theorem 219 *Suppose $G_s = (-1)^{l-n} D^{2(l-n)} \left((*\Lambda)^l\right)$ where Λ is the 1-dimensional hat function and n, l are integers such that $1 \leq n \leq l$.*

Then for given $\lambda > 0$, $G_s(\lambda x)$ is called a scaled extended B-spline basis function. The corresponding weight function is $w_\lambda(t) = 2\lambda a w\left(\frac{t}{2\lambda}\right)$ where w is the extended B-spline weight function with parameters n, l and $a = \frac{(2\pi)^{l/2}}{2^{2(l-n)+1}}$. Indeed, for w_λ we can choose $\kappa = n - 1$. Further

$$G_s(0) - G_s(\lambda x) \leq \lambda \left\|DG_s\right\|_\infty |x|, \quad x \in \mathbb{R}^1. \tag{5.61}$$

Finally, if $f_d \in X_w^0$ and $g_d(x) = f_d(2\lambda x)$, it follows that $g_d \in X_{w_\lambda}^0$ and $\left\|g_d\right\|_{w_\lambda,0} = \sqrt{a}\left\|f_d\right\|_{w,0}$.

Numerical results

$\boxed{\text{Smoother error on Data region:}}$ As the number of data points increases the error function of the smoother of f_d stabilizes: it has alternating positive and negative spikes with an overall mean of zero. The zeros of the error function have a period of 3/16: note there are 16 cells in the smoother grid.

As the number of data points increases the error of the smoother of the Riesz data function R_0 is observed to stabilize: it has a wide negative central spike at $x = 0$ surrounded by much smaller values.

$\boxed{\text{Absolute smoother error vs. Data density:}}$ As with the previous case we select $\rho = 10^{-6}$ and, using the 'spike filter' of the previous case, we obtain the four subplots displayed in Figure 5.6, each subplot being the superposition of 20 smoothers. The two upper subplots relate to the data function 5.60 and the lower subplots relate to the Riesz data function R_0.

The (blue) points above the actual errors in Figure 5.6 are the estimated upper bounds for the error given by the estimates 5.44 and 5.45. Clearly both theoretical bounds substantially underestimates the convergence rate. The instabilities associated with small numbers of data points eventually disappear as the number of data points increases. The smoother of R_0 is less stable than that of f_d.

5.10.2 Extended B-splines with $n = 2$

Since $n \geq 2$ we can use the error estimates of Theorem 210. We will use the same scaled (dilated) B-spline and data function that we used for interpolation in 2.6.2 and for Exact smoothing in Subsection 4.8.2.

The case $n = 2$, $l = 2$

The basis function is

$$G_{2,2}(x) = \frac{(\Lambda * \Lambda)(2x)}{(\Lambda * \Lambda)(0)} = \frac{3\sqrt{2\pi}}{2}(\Lambda * \Lambda)(2x),$$

with scaling factor $\lambda = 4$ and is such that $\operatorname{supp} G_{2,2} = [-1, 1]$ and $G_{2,2}(0) = 1$. To calculate $G_{2,2}$ we use the convenient formula

$$G_{2,2}(x) = (1 + x)^2 \Lambda(2x + 1) + \left(1 - 2x^2\right)\Lambda(2x) + (1 - x)^2 \Lambda(2x - 1),$$

and choose the bell-shaped data function

$$f_d = \delta_2^2 U \in X_w^0, \tag{5.62}$$

where

$$U(x) = \frac{e^{-k_{1,2}x^2}}{\delta_2^2\left(e^{-k_{1,2}x^2}\right)(0)} = \frac{e^{-k_{1,2}x^2}}{2\left(1 - e^{-4k_{1,2}}\right)}, \quad k_{1,2} = 0.3, \tag{5.63}$$

so that

$$k_G = \frac{\sqrt{12}}{(2\pi)^{1/4}}, \quad \left\|f_d\right\|_{w,0} = \sqrt[4]{72\pi}\frac{(k_{1,2})^{3/4}}{1 - e^{-4k_{1,2}}}, \quad k_{1,2} = 0.3, \quad s = 1, \quad h_G = \infty. \tag{5.64}$$

For the double rate convergence experiments the data function is $R_0 = (2\pi)^{-\frac{1}{2}} G_{2,2}$ where $R_0(0) = (2\pi)^{-\frac{1}{2}}$.

The next step is to calculate $\|f_d\|_{\infty,K}$ and $\|R_0\|_{\infty,K}$ where $K = [-1.5, 1.5]$. Clearly $\|R_0\|_{\infty,K} = R_0(0)$ and from 5.62 and 5.63, $\|f_d\|_{\infty,K} = f_d(0) = 1$. Thus

$$\|R_0\|_{\infty,K} = R_0(0) = (2\pi)^{-\frac{1}{2}}, \quad \|f_d\|_{\infty,K} = 1. \tag{5.65}$$

<u>Numerical results</u>

Smoother error on Data region: As the number of data points increases the smoother error of f_d is observed to consist of stable alternating positive and negative spikes with mean zero. The zeros of the error function have a period of 3/16: note there are 16 cells in the smoother grid.

As the number of data points increases the smoother error of the Riesz data function R_0 is observed to stabilize: it has a wide negative central spike at $x = 0$ surrounded by much smaller values.

Absolute smoother error vs. Data density: As in the previous cases we select $\rho = 10^{-6}$ and use the spike filter. The output consists of the four subplots displayed in Figure 5.7, each subplot being the superposition of 20 smoothers. The bottom subplots correspond to the Riesz data function.

The (blue) points above the actual errors in Figure 5.7 are the theoretical upper bounds for the smoother error given by the estimates 5.44 and 5.45.

Clearly for the data function f_d the theoretical error bound substantially underestimates the convergence rate. Regarding the data function R_0, the theoretical upper error bound for the data function R_0 is a rather better fit to the unfiltered errors. The instabilities associated with small numbers of data points eventually disappear as the number of data points increases.

It is interesting to note that the case of 8 smoother grid cells shown in Figure 5.8 is more stable and converges to a much smaller error value than the case of 16 smooth grid cells in Figure 5.7.

5.11 A numerical implementation of the Approximate smoother

5.11.1 The SmoothOperator software (freeware)

In this section we discuss a numerical implementation for the construction and solution of the zero order Approximate smoother matrix equation 5.13 i.e.

$$\left(N\rho R_{X',X'} + R^T_{X,X'} R_{X,X'}\right)\alpha' = R^T_{X,X'} y.$$

I called this software *SmoothOperator*. In Subsection 5.5.2 it was shown that the construction and solution of this matrix equation is a scalable process and thus worthy of numerical implementation. This software also implements the positive order version of this equation which is derived in Williams [13] but the tutorials concentrate on the zero order basis functions.

The algorithm has been implemented in **Matlab 6.0** with a GUI interface but has only been tested on Windows. *SmoothOperator* can be obtained by emailing the author. However there is a **short user document** (4 pages) and the potential user can read this document first to decide whether they want the software. The top-level directory of the software contains the file *read_me.txt* which can also be downloaded separately. A **full user manual** (82 pages) comes with the software. The main features of the full user manual are:

1 Tutorials and data experiments To learn about the system and the behavior of the algorithm, I have prepared three tutorials and five data experiments.

2 Context-sensitive help Each dialog box incorporates context-sensitive help which is invoked using the right mouse button. An F1 key facility could be implemented. The actual help text is contained in the text file **\Help\ContextHelpText.m** that can be easily edited.

3 Matlab diary facility When the system is started the Matlab diary facility is invoked. This means that most user information generated by *SmoothOperator* and written to the Matlab command line is also written to the text diary file. Each dialog box has a drop-down diary menu which allows the user to

view the diary file using Notepad. There is also a facility for you to choose another editor/browser. The file can also be emptied, if, for example, it gets too big. It can also be disabled.

4 Tools for viewing data files Before generating a smoother you can view the contents of the data file.

For ASCII delimited text data files, use the (slow) **View records** facility to display records and the file header, and then with this information you can use the high speed **Study records** facility to check the records and then obtain detailed information about single fields e.g. a histogram, and multiple fields e.g. correlation coefficients and scatter plots.

For binary **test data files** only the **View records** facility is needed. The parameters which generated the file can also be viewed using the **Make or View data** option.

5 The output data The output from the experiments and tutorials mentioned in point 1 above consists of well-documented Matlab one and two-dimensional plots, command line output and diary output. There is currently no file output, but this can be implemented on request.

6 Reading delimited text files A MEX C file allows ASCII delimited text files to be read very quickly. You can specify:
6.1 that the file be read in chunks of records, and not in one go.
6.2 The ids of the fields to be read. This means that the file can contain non-numeric fields.
6.3 Fields can be checked to ensure they are numeric.

The tutorials which create smoothers allow the data, smoothed data, and related functions to be viewed using scatter plots and plots along lines and planes.

7 Please note that there is no explicit suite of functions - application programmer interface or API - supplied by *SmoothOperator* for immediate use by the user. After reflecting on how I would create such an API, I decided that I lacked the experience to produce it, and that, anyhow, there were too many possibilities to anticipate. I would expect that possible users of this software, designed to be applied to perhaps millions of records, would need to familiarize themselves thoroughly with how this system works. I urge them to contact me and discuss their application.

5.11.2 Algorithms

The following three algorithms are used in the *SmoothOperator* software package to calculate the Approximate smoother.

Algorithm 1 uses data generated internally according to specifications supplied by the user using the interface. The smoothing parameter can be either specified or calculated using an error grid. This algorithm is **scalable**.

Algorithm 2 uses a 'small' subset of actual data to get an idea of a suitable smoothing parameter to use for the full data set. This algorithm is **not scalable**.

Algorithm 3 uses all the actual data. The smoothing parameter can be either specified or calculated using an *error grid*. This algorithm is **scalable**.

We will now explain these algorithms in more detail.

<u>Algorithm 1: using experimental data</u>

1. Generate the experimental data $[X, y]$. The independent data X is generated by uniformly distributed random numbers on X. The dependent data y is generated by uniformly perturbing an analytic data function g_{dat}.

2. Choose a smoothing grid X' whose boundary contains X. Choose an error grid X'_{err} which will be used to estimate the optimal smoothing parameter ρ.

3. Read the data $[X, y]$ and construct the matrix equation. If the matrix $G_{X,X'}$ is dense, the matrix $G^T_{X,X'} G_{X,X'}$ is constructed using a Matlab mex file (a compiled C file). If $G_{X,X'}$ is sparse the usual matrix multiplication is used.

4. Given a value for the smoothing parameter ρ we can solve the matrix equation and evaluate the smoother.

We want to estimate the value of ρ which minimizes the 'sum of squares' error between the smoother and the data function

$$\delta_1(\rho) = \sum_{x'' \in X'_{err}} \left(\sigma_\rho(x'') - g_{dat}(x'')\right)^2, \quad \rho > 0.$$

Empirical work indicates a standard shape for $\delta_1(\rho)$, namely decreasing from right to left, reaching a minimum and then increasing at a decreasing rate. To find the minimum we basically use the standard iterative algorithm of dividing and multiplying by a factor e.g. 10 and choosing the smallest value. The process is stopped when the percentage change of one or both of $\delta_1(\rho)$ and ρ are less than prescribed values.

Algorithm 2: using a 'test' subset of actual data

1. Perhaps based on results using Algorithm 1, choose an initial value for smoothing parameter ρ and choose a smoothing grid X'.
2. Step 2 of Algorithm 1. Denote the data by $[X, y]$ where $X = \left(x^{(i)}\right)$ and $y = (y_i)$.
3. This is the same as step 4 of Algorithm 2 except we now minimize

$$\delta_2(\rho) = \sum_{i=1}^{N} \left(\sigma_\rho\left(x^{(i)}\right) - y_i\right)^2, \quad \rho > 0,$$

because we do not have a data function.

Algorithm 3: using all the actual data

1. Choose a value for smoothing parameter ρ, based on experiments using Algorithm 2. Choose a smoothing grid X'.
2. Read all the data $[X, y]$ and construct the matrix equation. If the matrix $G_{X,X'}$ is dense, the matrix $G^T_{X,X'} G_{X,X'}$ is constructed using a Matlab mex file (a compiled C file). If $G_{X,X'}$ is sparse the usual multiplication is used.
3. Solve the matrix equation to obtain the basis function coefficients and evaluate the smoother at the desired points.

5.11.3 Features of the smoothing algorithm and its implementation

1 The Short user manual and the User manual contains a lot of detail regarding the *SmoothOperator* system and algorithms. So we will content ourselves here with just some key points.

2 Although the algorithm is scalable there can still be a problem with rapidly increasing memory usage as the grid size decreases and the dimension increases. The classical radial basis functions, such as the thin plate spline functions, have support everywhere. Hence the smoothing matrix is completely full. To significantly reduce this problem we do the following :
a) We use basis functions with bounded support.
b) We shrink the basis function support to the magnitude of the grid cells. This makes $G^T_{X,X'} G_{X,X'}$ a **very sparse** banded diagonal matrix, and $G_{X',X'}$ has a small number of non-zero diagonals. We say this basis function has *small support.*

3 Instead of using the memory devouring Matlab `repmat` function to calculate $G_{X,X'}$ and $G_{X',X'}$, we directly calculate the arguments of the Matlab function `sparse`. This function takes three arrays, namely the row ids, the column ids and the corresponding matrix elements, as well as the matrix dimensions, and converts them to the Matlab sparse internal representation. I must mention that although this is quick and space efficient, the algorithm is much more complicated than using `repmat`.
Matlab's sparse multiplication facility is then used to quickly and efficiently calculate $G^T_{X,X'} G_{X,X'}$.

4 This software has implemented the above techniques for the smoothing matrix, and a selection of basis functions is supplied.

5 The tutorials and exercises are based around the zero order **tensor product hat (triangle) function,** denoted by Λ. The hat function, also known as the triangle function, is used because of its simplicity and its analytic properties. The higher dimensional hat functions are defined as the tensor product of the one dimensional hat function : $\Lambda(x_1) = 1 - |x_1|$, when $|x_1| \leq 1$, and zero otherwise. This function has zero order so that the smoothing matrix simplifies to

$$\Psi_G = \rho N G_{X',X'} + G^T_{X,X'} G_{X,X'}, \quad G_{X',X'} = I, \tag{5.66}$$

with matrix equation

$$\Psi_G \alpha = G_{X',X} y.$$

Smoothers constructed from the hat function are continuous but not smooth. I have also included the **B-spline of order 3**, namely $\Lambda * \Lambda$, which is a basis function of order zero or one. This can be used when a smoother is required to be smooth i.e. continuously differentiable.

6 Note that this software was written before I embarked on the error analysis contained in this document so *SmoothOperator* concentrates on the Approximate smoother and the user cannot study the error of the Exact smoother or the Approximate smoother or compare the Approximate smoother with the Exact smoother.

5.11.4 An application - predictive modelling of forest cover type

In this section we demonstrate how the method developed in the previous sections can be used in data mining for predictive modelling. This application smooths **binary-valued** data - I was unable to obtain a large 'continuous-valued' data set for distribution on the web. The source of our data is the web site file:

http://kdd.ics.uci.edu/databases/covertype/covertype.html,

in the UCI KDD Archive, Information and Computer Science, University of California, Urvine.

The data gives the forest cover type in 30×30 meter cells as a function of the following cartographic parameters:

Id	Independent variable	Description
1	ELEVATION	Altitude above sea level
2	ASPECT	Azimuth
3	SLOPE	Inclination
4	HORIZ_HYDRO	Horizontal distance to water
5	VERT_HYDRO	Vertical distance to water
6	HORIZ_ROAD	Horizontal distance to roadways
7	HILL_SHADE_9	Hill shade at 9am
8	HILL_SHADE_12	Hill shade at noon
9	HILL_SHADE_15	Hill shade at 3pm
10	HORIZ_FIRE	Horizontal distance to fire points

TABLE 5.3.

Forest cover type is the dependent variable y and it takes on one of seven values:

Forest cover type	Id
Spruce fir	1
Lodge-pole pine	2
Ponderosa pine	3
Cottonwood/Willow	4
Aspen	5
Douglas fir	6
Krummholtz	7

TABLE 5.4.

The data file

In this study we will use the data to train a model predicting on the **presence or absence of the Ponderosa pine forest cover** (id = 3), but the results will be similar for the other forest types. To this end we have created from the full web site file a file called `\UserData\forest_1_to_10_pondpin.dat` which contains the ten independent variables of Table 5.3 and then a binary dependent variable derived from the variable **Ponderosa pine** of Table 5.4. This variable is 1 if the cover is Ponderosa pine and zero otherwise.

Methodology

We recommend you first use the user interface of the *SmoothOperator* software to construct artificial data sets to understand the behavior of the smoother and run the experiments to get a feel for the influence of the parameters.

1 We chose a small subset of the forest-cover data and selected various smoothing parameters and smoothing grid sizes to study the performance of the smoother using plots and the value of the error.

Note that *two subsets of data* could be used here, often called the *training set* and the *test set*. The training set would be used to calculate the smoother for the initial value of the smoothing coefficient, and then the test set could be used to determine the smoothing coefficient which minimizes the least squares error.

2 Having chosen our parameters we run the smoother program on the full data set.

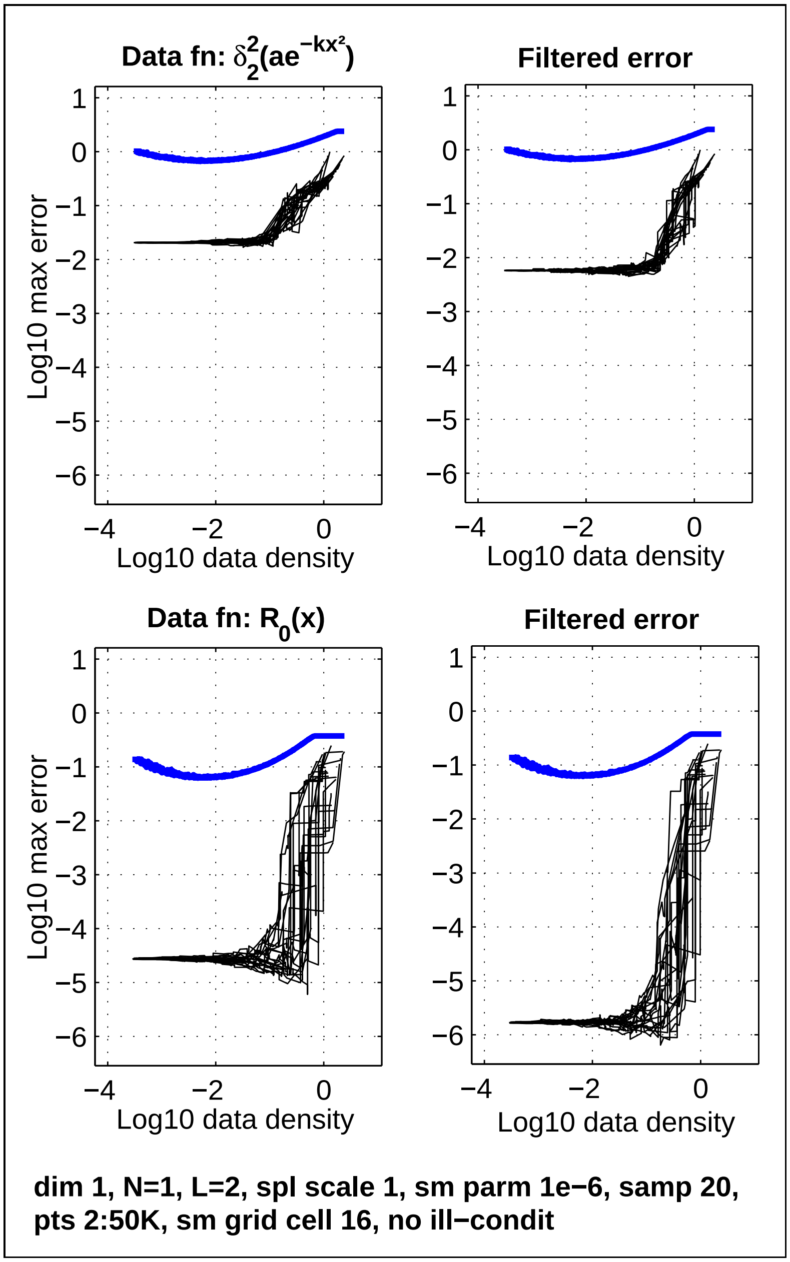

FIGURE 5.6. Error of Approximate smoother vs data density.

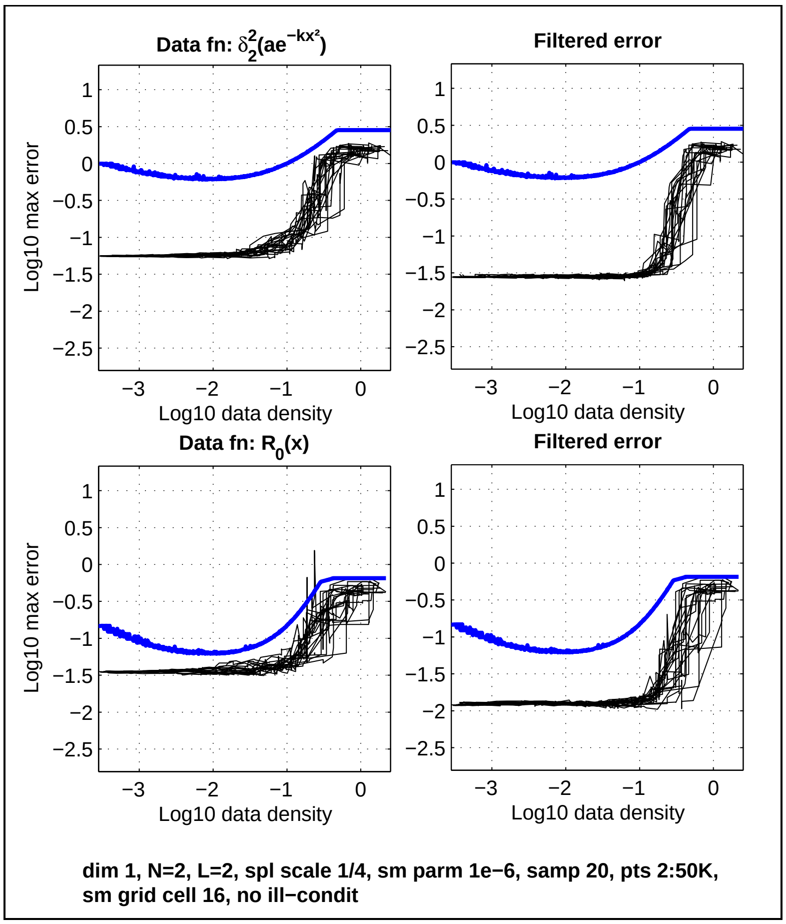

FIGURE 5.7. Error of Approx smoother vs data density: 16 smth grid cells.

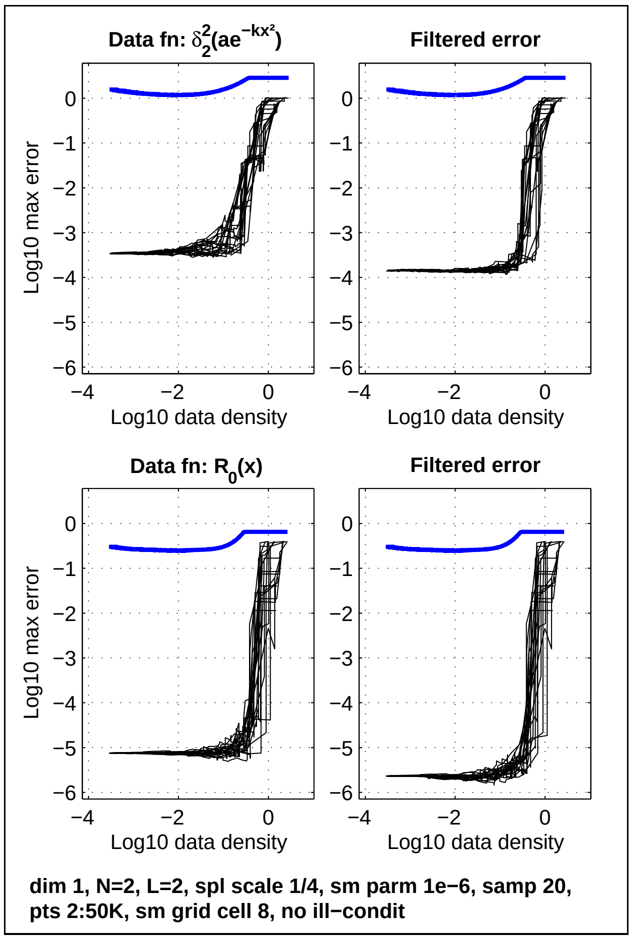

FIGURE 5.8. Error of Approx smoother vs data density: 8 smth grid cells.

Appendix A

Basic notation, definitions and symbols

A.1 Basic function spaces

Definition 220 ***Basic function spaces***

All spaces below consist of ***complex-valued*** *functions.*

- $P_0 = \{0\}$. *For* $n \geq 0$, P_n *denotes the polynomials of (total) order at most* n *i.e. degree at most* $n-1$ *when* $n \geq 1$. *These polynomials have the form* $\sum\limits_{|\alpha|<n} a_\alpha \xi^\alpha$, *where* $a_\alpha \in \mathbb{C}$ *and* $\xi \in \mathbb{R}^d$.

- $C^{(0)}$ *is the space of continuous functions.*

- $C_B^{(0)}$ *is the space of bounded continuous functions.*

- $C_{BP}^{(0)}$ *is the space of continuous functions bounded by a polynomial.*

- $C^{(m)} = \left\{f \in C^{(0)} : D^\alpha f \in C^{(0)},\ when\ |\alpha| = m\right\}$.

- $C_B^{(m)} = \left\{f \in C_B^{(0)} : D^\alpha f \in C_B^{(0)},\ when\ |\alpha| \leq m\right\}$.

- $C_{BP}^{(m)} = \left\{f \in C_{BP}^{(0)} : D^\alpha f \in C_{BP}^{(0)},\ when\ |\alpha| \leq m\right\}$.

- $C^\infty = \bigcap\limits_{m \geq 0} C^{(m)}$; $C_B^\infty = \bigcap\limits_{m \geq 0} C_B^{(m)}$; $C_{BP}^\infty = \bigcap\limits_{m \geq 0} C_{BP}^{(m)}$.

- C_0^∞ *is the space of* C^∞ *functions that have compact support.*

  These are the test functions for the space of distributions.

- S *is the* C^∞ *space of rapidly decreasing functions. These are the test functions for the tempered distributions of Definition 223 below.*

- *For a bounded continuous function on* $\mathbb{R}^d$, $\|f\|_\infty = \sup\limits_{x \in \mathbb{R}^d} |f(x)|$.

- L_{loc}^1 *is the space of measurable functions which are absolutely integrable on any compact set i.e. any closed, bounded set.*

- L^1 *is the complete normed vector space of measurable functions* f *such that* $\int |f| < \infty$. *Norm is* $\|f\|_1 = \int |f|$.

- L^2 *is the Hilbert space of measurable functions* f *such that* $\int |f|^2 < \infty$. *Norm is* $\|f\|_2 = \left(\int |f|^2\right)^{1/2}$ *and inner product is* $(f,g)_2 = \left(\int f\overline{g}\right)^{1/2}$.
- L^∞ *is the complete normed vector space of essentially bounded functions. The norm is* $\|f\|_\infty = \operatorname*{essup}_{x\in\mathbb{R}^d} |f(x)|$.

A.2 Multi-index and vector notation

Definition 221 ***Notation***

1. *Multi-indexes are vectors with non-negative integer components.*
   Let $x = (x_1, x_2, \ldots, x_d)$ *and* $y = (y_1, y_2, \ldots, y_d)$.
   Suppose $\tilde{}$ *is one of the binary operations* $<$, $\leq$, $=$, $>$, $\geq$.
   Write $\beta\tilde{}a$ *if* $\beta_i\tilde{}\alpha_i$ *for all* i.
   For $x \in \mathbb{R}$ *write* $\beta\tilde{}x$ *if* $\beta_i\tilde{}x$ *for all* i.
   The component-wise product is denoted $x.y = (x_i y_i)$. *Component-wise division is denoted* $x/y = (x_i/y_i)$ *or* $\frac{x}{y} = \left(\frac{x_i}{y_i}\right)$.
   Denote $\mathbf{1} = (1,1,1,\ldots,1)$ *and let* $\{\mathbf{e}_k\}_{k=1}^d$ *be the canonical (or standard) basis of* $\mathbb{R}^d$.
2. *Denote* $|\alpha| = \sum_{i=1}^d \alpha_i$. *The* $D^\alpha f(x)$ *is the derivative of the function* f *of degree* α
   $$D^0 f(x) = f(x), \quad D^\alpha f(x) = \frac{D^{|\alpha|} f(x_1, x_2, \ldots, x_d)}{D_1^{\alpha_1} x_1 D_2^{\alpha_2} x_2 \ldots D_d^{\alpha_d} x_d}.$$
3. $$\begin{array}{ll} \textit{monomial} & x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_d^{\alpha_d}, \\ \textit{factorial} & \alpha! = \alpha_1!\alpha_2!\ldots\alpha_d! \textit{ and } 0! = 1, \\ \textit{binomial} & \dbinom{\alpha}{\beta} = \dfrac{\alpha!}{(\alpha-\beta)!\beta!}, \quad \textit{if } \beta \leq \alpha. \end{array}$$
4. *The inequality* $|x^\alpha| \leq |x|^{|\alpha|}$ *is used often.*
5. *Important identities are*
   $$(x,y)^k = \sum_{|\alpha|=k} \frac{k!}{\alpha!} x^\alpha y^\alpha, \quad |x|^{2k} = \sum_{|\alpha|=k} \frac{k!}{\alpha!} x^{2\alpha}. \tag{A.1}$$
6. *If* $\binom{k}{\alpha} = \frac{k!}{\alpha!(k-|\alpha|!)}$ *then*
   $$\left(1 + |x|^2\right)^k = \sum_{|\alpha|\leq k} \binom{k}{\alpha} x^{2\alpha}.$$
7. *The binomial expansion is*
   $$(x+y)^\alpha = \sum_{\beta\leq\alpha} \binom{\alpha}{\beta} x^\beta y^{\alpha-\beta},$$
   and a consequence is
   $$\sum_{\beta\leq\alpha} \binom{\alpha}{\beta} = 2^{|\alpha|}.$$

8. *A useful identity is*

$$\sum_{|\alpha|=k} 1 = \binom{d+k-1}{k}.$$

9. *Leibniz's rule is*

$$D^{\alpha}(uv) = \sum_{\beta \leq \alpha} \binom{\alpha}{\beta} D^{\beta} u D^{\alpha-\beta} v.$$

A.3 Topology

Definition 222 ***Topology on*** $\mathbb{R}^d$

- *The Euclidean norm is denoted by* $|x|$ *and the inner product by* xy *or* (x, y).

  Other norms are $|x|_1 = \sum_{i=1}^{d} |x_i|$ *and* $|x|_\infty = \max_i |x_i|$.
- *The (open) ball* $B(x; r) = \{y : |x - y| < r\}$.
- $\varepsilon-$*neighborhood of a set: for* $\varepsilon > 0$ *the* $\varepsilon-$*neighborhood of a set* S *is* $S_\varepsilon = \bigcup_{x \in S} B(x; \varepsilon)$.
- $R(a, b) = \{x : a < x < b\}$ *will denote an open rectangle with left and right points* a *and* b.
- $R[a, b) = \{x : a \leq x < b\}$ *will denote the partially open rectangle with left and right points* a *and* b.

A.4 Tempered distributions

Definition 223 ***Tempered distributions (or generalized functions of slow growth)***

1. S *is the space of rapidly decreasing* C^∞ *functions We endow* S *with the topology defined using the countable set of seminorms*

$$\|(1 + |\cdot|)^n D^\alpha \psi\|_\infty, \quad \psi \in S,\, n = 0, 1, 2, \ldots;\ \alpha \geq 0.$$

2. S' *denotes the space of tempered distributions or generalized functions of slow growth. It is the set of all continuous linear functionals on* S *under the seminorm topology of part 1.*
3. *If* $f \in S'$ *and* $\phi \in S$ *then* $[f, \phi] \in \mathbb{C}$ *will represent the action of* f *on the test function* ϕ. *The functions in* S *are called the test functions of* S'.

A.5 Fourier Transforms

Definition 224 ***Fourier and Inverse Fourier transforms on*** S ***and*** S'.

1. *This document uses the two Fourier transform notations*

$$F[f] = \widehat{f}(\xi) = (2\pi)^{-d/2} \int\limits_{\mathbb{R}^d} e^{-ix\xi} f(x) dx, \quad f \in S,$$

so that $F : S \to S$. *The inverse Fourier transform* $F^{-1} : S \to S$ *is*

$$F^{-1}[f] = \overset{\vee}{f}(\xi) = (2\pi)^{-d/2} \int\limits_{\mathbb{R}^d} e^{ix\xi} f(x) dx, \quad f \in S.$$

2. The Fourier/Inverse Fourier transforms are extended to mappings from S' to S' by

$$\left[\widehat{f},\phi\right]=\left[f,\overset{\vee}{\phi}\right];\qquad\left[\overset{\vee}{f},\phi\right]=\left[f,\overset{\vee}{\phi}\right],\quad f\in S',\phi\in S.$$

Some important properties are:

1. $\overset{\vee}{f}(\xi)=\widehat{f}(-\xi)$, $\overline{\widehat{f}}(\xi)=\overset{\vee}{\overline{f}}(\xi)$ and $\widehat{\widehat{f}}(\xi)=f(-\xi)$.
2. $(f(x-a))^{\wedge}=e^{-ia\xi}\widehat{f},\quad a\in\mathbb{R}^d$.
3. $\left(e^{iax}\right)^{\wedge}=(2\pi)^{d/2}\,\delta(\xi+a),\quad a\in\mathbb{R}^d$.
4. $D^\alpha\left(\widehat{f}\right)=(-i)^{|\alpha|}\,\widehat{x^\alpha f}$.
5. $\widehat{x^\alpha f}=i^{|\alpha|}D^\alpha\widehat{f}$.
6. $\widehat{D^\alpha f}=i^{|\alpha|}\xi^\alpha\widehat{f}=(i\xi)^\alpha\,\widehat{f}$.
7. $\xi^\alpha\widehat{f}=(-i)^{|\alpha|}\,\widehat{D^\alpha f}$.
8. $\widehat{\delta}=(2\pi)^{-d/2}$ and $\widehat{1}=(2\pi)^{d/2}\,\delta$.
9. $\widehat{x^\alpha}=(2\pi)^{d/2}\,(iD)^\alpha\,\delta$ and $\widehat{D^\alpha\delta}=(2\pi)^{-d/2}\,(-i)^{|\alpha|}\,\xi^\alpha$.
10. If p is a polynomial then $\widehat{p}=(2\pi)^{d/2}\,p(iD)\delta$ and $\widehat{p(D)\,f}=p(-i\xi)\,\widehat{f}$.

A.6 Convolutions

Definition 225 ***Convolution***
If $f\in\mathbb{C}_{BP}^{(0)}$ and $\phi\in S$ then the convolution of f and ϕ is denoted by $f*\phi$ or $\phi*f$ and is defined by

$$(\phi*f)(x)=(2\pi)^{-d/2}\int f(x-y)\,\phi(y)\,dy=(2\pi)^{-d/2}\int f(y)\,\phi(x-y)\,dy.$$

This definition is extended to $f\in S'$ by the formulas

$$f*\phi=\left(\widehat{\phi}\widehat{f}\right)^{\vee}=(2\pi)^{-d/2}\left[f_y,\phi(\cdot-y)\right],\quad\phi\in S.\tag{A.2}$$

A.7 Taylor's expansion with integral remainder

Suppose $u\in C^{(n)}\left(\mathbb{R}^d\right)$ for some $n\geq 1$. Then the Taylor series expansion about z is given by

$$u(z+b)=\sum_{|\beta|<n}\frac{b^\beta}{\beta!}\left(D^\beta u\right)(z)+\left(\mathcal{R}_n u\right)(z,b),$$

where the integral remainder term

$$\begin{aligned}\left(\mathcal{R}_n u\right)(z,b)&=n\sum_{|\beta|=n}\frac{b^\beta}{\beta!}\int_0^1 t^{n-1}(D^\beta u)(z+(1-t)\,b)dt\\&=n\sum_{|\beta|=n}\frac{b^\beta}{\beta!}\int_0^1(1-t)^{n-1}\,(D^\beta u)(z+tb)dt,\end{aligned}$$

satisfies the estimate (proved using A.1)

$$\left|\left(\mathcal{R}_n u\right)(z,b)\right|\leq\frac{d^{\frac{n}{2}}\,|b|^n}{n!}\max_{\substack{|\beta|=n\\ t\in[z,z+b]}}\left|D^\beta u(t)\right|.\tag{A.3}$$

References

[1] R. A. Adams, *Sobolev spaces*, Academic Press, New York, 1975.

[2] J. Duchon, *Splines minimising rotation-invariant seminorms in sobolev spaces*, Constructive theory of functions of several variables (Schempp W. and Zeller K., eds.), Lecture Notes in Mathematics, vol. 571, Springer-Verlag (Berlin), 1977, pp. 85–100.

[3] N. Dyn, *Interpolation and approximation by radial and related functions*, Approximation Theory VI (Chui C., Schumaker L., and Ward J., eds.), Academic Press, 1989, pp. 211–234.

[4] J. Garke and M. Griebel, *Semi-supervised learning with sparse grids*, Proceedings of the 22nd International Conference on Machine Learning, 2005.

[5] K. Hollig, *Finite element methods with b-splines*, Frontiers in applied mathematics, vol. 3, Society Industrial and Applied Mathematics, Philadelphia, PA, 2003.

[6] W. Light and H. Wayne, *Error estimates for approximation by radial basis functions.*, Approximation theory, wavelets and applications (Maratea, 1994), NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., vol. 454, Kluwer Acad. Publ., Dordrecht, 1995, pp. 215–246.

[7] ———, *On power functions and error estimates for radial basis function interpolation*, J. Approx. Th. **92** (1998), no. 2, 245–266.

[8] ———, *Spaces of distributions, interpolation by translates of a basis function and error estimates*, Numer. Math. **81** (1999), no. 3, 415–450.

[9] W. Magnus, F. Oberhettinger, and R. Soni, *Formulas and theorems for the special functions of mathematical physics*, 3rd ed., Springer-Verlag, Berlin Heidelberg New York, 1966.

[10] P. Malliavin, H. Airault, L. Kay, and G. Letac, *Integration and probability*, Springer-Verlag, Berlin, 1995.

[11] B. Petersen, *Introduction to the fourier transform and pseudo-differential operators*, Pitman, 1983.

[12] V. Vladimirov, *Equations of mathematical physics*, Marcel Dekker, New York, 1971.

[13] P. Y. Williams, *A weight function theory of positive order basis function interpolants and smoothers.*, Prepublication: http://au.arxiv.org/.

[14] K. Yosida, *Functional analysis*, Springer-Verlag, Berlin, 1958.